MEASURING THE DISTRIBUTION OF SELF-REPORTED HEALTH

By

Robert Andrew Allison

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

December, 1999

Nashville, Tennessee

Approved:	Date:
	11/29/99
	11/29/99
	11/29/99
	11/30/99
	12/1/99

ECONOMICS

MEASURING THE DISTRIBUTION OF SELF-REPORTED HEALTH

ROBERT ANDREW ALLISON

Dissertation under the direction of
Professor James E. Foster

Many questions in health policy require an understanding of the distribution of health across a given population and how it changes as a result of policy interventions. Since objective data on individual health status are often unavailable or incomplete, especially for populations with very low mortality, increasing use has been made of Self-Reported Health Status (SRHS) data, which record people's own perceptions of their health status. The qualitative or categorical nature of SRHS data prevents the straightforward use of traditional tools of distributional analysis such as the Lorenz curve and the mean. This dissertation explores a number of alternatives for measuring the distribution of self-reported health.

Chapter II examines a prominent method introduced in Wagstaff and van Doorslaer (1994) that employs numerically scaled SRHS data to measure socioeconomic inequalities in health. The chapter demonstrates several potential limitations of this method, especially when applied in an international context.

Chapter III presents a new methodology for evaluating qualitative distributions that generates results that do not depend on the numeric scaling assigned to the categories. A partial inequality ordering is defined that indicates when a distribution is

more "spread out" than another; a second partial ordering (first order stochastic dominance) is used to indicate when the overall health level rises. It has been argued that reasonable scalings of SRHS data are concave, emphasizing distances between lesser health categories. Chapter IV introduces partial distributional orderings (using second order dominance) that apply to all reasonable scalings of the data, so that results do not depend on which concave scale is used.

The methods introduced in chapters III and IV are illustrated using SRHS data from the 1990-1994 *National Health Interview Survey State Data Files*. Results indicate that southern states generally have less overall health, higher health poverty rates, and a more unequal distribution of health than states in other regions.

The dissertation concludes in chapter V with a brief empirical analysis that identifies key determinants of interstate differences in health and demonstrates how these results can be used to evaluate the mechanism for distributing federal aid to states via the Medicaid program.

Approved _____ Date Nov 22 1995

To Susan and the children

ACKNOWLEDGEMENTS

I wish to thank Dr. James Foster, my advisor and coauthor, for inspiration in the classroom and throughout the research phase, and for allowing and encouraging me to pursue this topic. I also thank Dr. Kathryn Anderson for her willingness and ability to help with this research project. I wish to thank Dr. Len Nichols for helping to inspire my interest in the study of economics. I thank my parents, Dr. and Mrs. Robert F. Allison, and my wife's parents, Rev. and Mrs. Frank R. Shell, for their prayers and for their support, which have been instrumental in helping me to complete this project. A special note of appreciation is due my father, Dr. Robert F. Allison, for the wisdom and intellect he has shared with me over the years.

I owe a debt of gratitude to my wife, Susan Shell Allison, for her invaluable commitment to me and to our children while I worked on this project. Finally, I give full credit to my Lord, Jesus Christ, for providing at every stage the means and the motivation to complete this Ph.D.

Financial Support

This research was supported in part by Dissertation Fellowship Award no. 30-P-91026/4-01 from the Health Care Financing Administration.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

Chapter

I. INTRODUCTION 1

II. THE USE OF QUALITATIVE HEALTH DATA TO CALCULATE THE HEALTH CONCENTRATION INDEX 7

1. Introduction 7
2. Interpreting the CI in an ideal empirical setting 10
3. Interpreting the CI when qualitative health data are used 19
4. Evaluating the concentration index using real-world SRHS data 26
5. Conclusion 38

III. MEASURING HEALTH INEQUALITY USING QUALITATIVE DATA 41
By R. Andrew Allison and James E. Foster

1. Introduction 41
2. Problems with the mean 45
3. A median-based approach 53
4. Health over the life cycle 61
5. Health across states 68
6. Conclusion 74

IV. THE DISTRIBUTION OF SELF-REPORTED HEALTH IN THE UNITED STATES: A RESTRICTED SCALE APPROACH 78

1. Introduction 78
2. Restricting the SRHS Scale 82
3. Second order distributional comparisons 85
4. Results 96
5. Conclusion 102

V. EXPLAINING BETWEEN-STATE VARIATION IN SELF-REPORTED HEALTH STATUS 104

1. Introduction 104
2. Identifying health needs among the states 108
3. Specification 110
4. Results 114
5. Discussion 118

Appendices

A. CHAPTER III PROOFS 121

B. CHAPTER IV PROOFS 126

BIBLIOGRAPHY 135

LIST OF TABLES

Table		Page
1.	CI Rankings Under Different Scalings of SRHS	29
2.	CI Rankings With Various SRHS Categories Combined	33
3.	SRHS Distributional Indicators for Females and Males Using the Linear Scale	65
4.	SRHS Distributional Indicators for the 25 Largest States Using the Linear Scale	70
5.	First Order Dominance in SRHS Among the 25 Largest States	72
6.	S Dominance in SRHS Among the 25 Largest States	73
7.	Second Order Comparisons of Overall SRHS Among the 25 Largest States	98
8.	Second Order S Inequality Comparisons of SRHS Among the 25 Largest States	100
9.	Robust Poverty Gap Comparisons of SRHS Among the 25 Largest States	101
10.	Cumulative Percentages at Each Level of SRHS in the 25 Largest States	112
11.	Effects of State Characteristics on Rates of Health Poverty (OLS)	115

LIST OF FIGURES

Figure		Page
1.	The Health Concentration Curve for Two Hypothetical Populations	12
2.	Plot of Health Against Income Rank for Two Randomly-Generated Distributions	18
3.	Mean Health Comparison of Men and Women, Selected Ages, Using Two Different Scales	47
4.	Cumulative Distribution Functions for x = (4,5,7,5,4) and y=(3,4,11,4,3)	54
5.	Deriving the S-Curve From the CDF	56
6.	S-Curves for x and y	56
7.	Variance in SRHS for Females	63
8.	Cumulative Probabilities at Each SRHS Category for Selected Ages	66
9.	Mean Ordering of x and y Under Two Different Scales	81
10.	Various Levels of Health in Relation to the Minimum Threshold Where Death Occurs	84
11.	Second Order S Inequality	93
12.	A Concave Transformation of Reference Scale c	126

CHAPTER I

INTRODUCTION

Public debate of health issues may be driven by any number of policy objectives. For many years the literature has emphasized of health equity objectives.[1] Tobin (1970), for example, distinguishes between the tolerance of Americans for income inequities that might result from individual choice, and their observed intolerance for inequities in the distribution of other commodities that might be necessary for "life, health, and citizenship." The literature displays a wide range of views as to the appropriate definition of health equity.[2] Muurinen and Le Grand (1985) suggest that equity is obtained through equality of access to health, while Culyer and Wagstaff (1993) define health equity as equality of health. Sen (1992) places health equity in a more comprehensive framework by characterizing the freedom from avoidable disease as one of many dimensions, or "spaces," in which individual disparities should be measured.[3] Sen's focus on avoidable disease is echoed by Kakwani, Wagstaff and van Doorslaer

---

[1]Others have emphasized society's interest in ensuring equitable access to health care. Of note are the views expressed by the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research in Securing Access to Health Care, Volume 1: Report (1983). The Commission discerned "in our country's traditional commitment to fairness an ethical obligation on the part of society to ensure that all Americans have access to an adequate level of care without the imposition of excessive burdens."

[2]Wagstaff and Van Doorslaer (1989) and Culyer and Wagstaff (1993) provide helpful reviews of this literature.

[3]Other dimensions might include adequate nourishment or having self-respect. See also Foster and Sen (1997) for a more complete discussion of this approach.

(1997), who criticize the inclusion in social indices of health equity those variations in health that are the direct result of unavoidable "biological influences" such as age and gender. Taken to an extreme, the focus on avoidable health inequality can result in a focus on specific causes or covariates of health differences. For example, Wagstaff, Paci, and van Doorslaer (1991) note that "what is interesting -- and indeed worrying -- about inequalities in health is not that they exist, but that they mirror inequalities in socioeconomic status."[4]

Two of the definitions of health equity described above receive further attention in this thesis: a broad definition of health equity as univariate health inequality, espoused by Culyer and Wagstaff (1993); and, a narrower definition of health equity as bivariate health equality, e.g., Wagstaff, Paci and Van Doorslaer's (1991) emphasis on equality in health across socioeconomic groups. A measurement strategy associated with the broader univariate equity agenda entails tracking the distribution of health, or some function of the distribution of health, in the populations that are to be compared, possibly controlling for certain specific sources of "unavoidable" variation. For example, one could follow Le Grand and Rabin (1986), who calculate the dispersion in age at death among males and females in Britain from 1931 to 1983, adjusting for differences in the age composition of each year. Le Grand and Rabin's objective was, obviously, to measure time trends in health equity in Britain. Alternatively, a cross-sectional analysis could be used to determine the (sub-) population most in need of inequality-reducing activities, though it might be difficult to use this kind of analysis to design or select

[4]P. 546.

specific interventions without also investigating the principle causes of variation in health.

A measurement strategy designed for the narrower bivariate equity objective would be to calculate differences in health status across population sub-groups, where the sub-groups are defined according to a specific individual characteristic, such as gender, age, or socioeconomic status. For example, one might calculate between-group inequality, which is a function of the difference between the mean level of health in each population sub-group. The level of health in each sub-group could be adjusted to remove variation due to factors that affect group identity as well as health status. For example, when measuring differences in health across genders, one might choose to adjust for differences in the age structure of males and females. Results from a comparison of health across genders might be used to determine which population was most in need of interventions designed to help equalize the health status of same-age males and females.

Since The Black Report (1982) analyzed mortality differentials across occupational classes, the most prevalent measurement strategy has been of the second type: comparing, or ranking, populations according to the degree of between-group inequalities found within them, specifically those inequalities attributable to socioeconomic status.[5] Even this bivariate approach, though, requires an understanding of the attributes of the distribution of health. Chapter II, for example, illustrates how one

[5]Wagstaff and van Doorslaer (1998a). Socioeconomic status could be defined by educational attainment, income level, wealth, or any number of other individual characteristics that impart status in a particular culture. Indeed, there is disagreement in the literature as to which particular co-variate best reflects socioeconomic status. See for example Kunst et al (1995) who use educational experience to identify social rank.

measure of socioeconomic inequality in health, the concentration index, depends crucially on the relative degree of dispersion in the distribution of health (see equation II.10). Measuring the distribution of health requires individual health status data. Since objective data on individual health status are often unavailable or incomplete, especially for populations with very low mortality, increasing use has been made of Self-Reported Health Status (SRHS) data, which record people's own perceptions of their health status. Yet SRHS data is qualitative, making it difficult to characterize basic attributes of the distribution. Two common approaches to the use of SRHS data to measure the distribution of health are: (i) to measure the prevalence of health poverty, e.g., the percentage of a group or population in the lowest one or two SRHS categories; and (ii) to calculate standard distributional indicators, such as the mean and variance, by assigning a particular scale to the data, such as the linear "1,2,3,4,5" scale. Both of these approaches require subjective judgements by the analyst that may limit results. Measuring health poverty entails the selection of a poverty threshold, so that results may depend on the choice of threshold. Similarly, distributional indicators calculated with numerically scaled SRHS data are functions of the selected scale, and may not hold when alternative scales are employed.

This dissertation explores a number of approaches for measuring the distribution of self-reported health. Chapter II examines a prominent method for measuring socioeconomic inequalities in health that was introduced in Wagstaff and van Doorslaer (1994). Their approach is to use SRHS data to calculate the health concentration index, which measures the rate at which health accumulates as socioeconomic rank increases. This chapter demonstrates a number of weaknesses in the method that may undermine its

intended use in international comparisons of socioeconomic inequality in health.

Chapter III presents a new methodology for evaluating qualitative distributions. A partial inequality ordering is defined that indicates when a distribution is more "spread out" than another; a second partial ordering (first order stochastic dominance) is used to indicate when the overall health level rises. Both relations are applicable to qualitative data, such as SRHS, in that results do not depend on the numeric scaling assigned to the categories. The approach is illustrated using SRHS data from the *National Health Interview Survey* and *NHIS State Data Files* for the 1990-1994 period, focusing on the distribution of SRHS within ages and states. Results indicate that southern states tend to have less overall health, higher health poverty rates, and a more unequal distribution of health than states in other regions.

While the distributional measures introduced in chapter III apply regardless of how the qualitative data are scaled, the measures introduced in chapter IV apply only to scales that are more concave than some reference scale. An argument can be made that convex scalings of SRHS data that emphasize distances between upper health categories are unreasonable. If so, the standard "1,2,3,4,5" scale can be used as reference scale to generate distributional comparisons that apply to all reasonable (concave) scalings of the SRHS data. The patterns of interstate disparities in aggregate health, health poverty and health inequality that are identified in chapter III are shown to persist over the set of additional comparisons made possible by the restricted-scale measurement approach in chapter IV.

Chapter V presents an example of how SRHS data might be used to support an explanation of differences in the distribution of health across groups or populations. In

this chapter, SRHS data from the *NHIS State Data Files* is used as the dependent variable in a series of regressions which explain interstate differences in health in the early 1990s. The chapter also explains how the analytic results might be used to evaluate the distribution of federal aid to states via the Medicaid program. The question addressed is whether the definition of health needs implicit in Medicaid eligibility rules is consistent with the health needs identified in the empirical analysis. In particular, do the primary determinants of Medicaid eligibility coincide with the key determinants of between-state differences in health needs across states? If not, then it is likely that Federal Medicaid payments will not be concentrated among states with the greatest health needs. The results suggest that at least one demographic characteristic that is a significant explanator of interstate differences in health -- educational attainment – is not reflected in Medicaid eligibility rules.

CHAPTER II

THE USE OF QUALITATIVE HEALTH DATA TO CALCULATE THE HEALTH CONCENTRATION INDEX

1. Introduction

There has been a good deal of interest recently in measuring and explaining the distribution of health in a population.[1] Numerous authors have, for example, attempted to measure differences in health associated with socioeconomic status, which is typically represented by income, educational attainment, or occupation. Among these efforts, the methodology generally attributable to Wagstaff and van Doorslaer (heretofore 'WvD'), and developed in Wagstaff, Paci, and van Doorslaer (1991), Wagstaff and van Doorslaer (1994), and Kakwani, Wagstaff, and van Doorslaer (1997), has received particular attention. The WvD methodology is designed to rank countries according to the rate at which health accumulates with socioeconomic rank. The inequality measure that WvD employ, the health concentration index (CI), is adapted from a technique developed by Kakwani (1977) to measure tax progressivity, and is basically a measure of health progressivity. The CI can be easily interpreted numerically and graphically, and is an intuitive measure of social disparities in health (see section 2 below).[2]

Since the CI measures the accumulation of health as social rank increases,

---

[1]See for example Measuring Health Globally (forthcoming).

[2]See Wagstaff, Paci, and van Doorslaer (1991) for a more complete discussion of the CI's attributes.

numerically scaled health data are required. To obtain such data, WvD assign numeric approximations to categorical self-reported health status (SRHS) data. It is not uncommon for researchers to assign numeric scores to qualitative data of this type in order to apply analytic techniques that require numeric data. In many cases, however, the dependence of the analytic results on these numeric assignments, or scales, is under-emphasized, and a sensitivity analysis using alternative scales is rare.

One strategy for applying a numeric scale to SRHS data is to assign unique numeric values to each categorical response. Along with the category frequencies obtained from survey data for some population, direct scaling of SRHS categories completely determines the attributes of the population's distribution of health. With this approach, the attributes of the distribution of health are free to vary across populations as the categorical frequencies fluctuate.[3] WvD employ an alternative strategy, deriving a numeric scale for SRHS data by fitting the observed frequencies in each category to an assigned continuous distribution. Unlike the direct scaling approach, WvD's method allows the numeric scale to vary across populations. As a result, the distances between, say, poor and fair health would be different in two compared distributions if the proportion of each group in the two categories was not identical; their (assigned) continuous distributions would, of course, look exactly the same. Presuming that we know too little about the distribution of health to make such an assignment, it is reasonable to question whether WvD's method is useful as a simplifying assumption. In particular, does the WvD method facilitate measurement of socioeconomic differences in

[3]A variant of the direct scaling approach is presented in chapters III and IV.

health without confounding those results? Several issues arise from this principle question. After assigning each compared population the same distribution of health, what does the CI really measure? How robust are CI comparisons to WvD's distributional assignment? What are the implications of allowing the scale assignments to vary across compared populations?

This chapter presents a review of WvD's use of SRHS data in the computation of the CI by characterizing the information imparted by CI rankings under various assumptions regarding the distribution of health. The chapter demonstrates how the ability of the CI to accurately rank populations in terms of the degree of socioeconomic inequality in health within them critically depends on the nature of the health data used. The chapter characterizes the effect that WvD's method of assigning SRHS scale values has on resulting CI rankings, and the potential challenges associated with using SRHS data to make international CI comparisons.

The remainder of this chapter is organized as follows. Section 2 introduces and defines the CI and interprets CI rankings in an ideal (and hypothetical) empirical setting where continuous and numerically-scaled data on individual health status is available. Evaluating the CI in this ideal setting helps to distinguish between the limitations of the CI that are inherent in the measure itself, and those that result from the use of less informative data. In section 3, WvD's scaling technique is introduced and the CI is reinterpreted in the context of an assigned distribution of health. This analysis demonstrates that if WvD's distributional assignment is a simplifying assumption, and the attributes of the distribution of health are not fully known, then the CI is best described as a sale-dependent measure of how uniformly health increases with income.

Section 4 illustrates the dependence of CI rankings on WvD's scaling methodology using data from the 1994 *National Health Interview Survey-State Data Files* (*NHIS-SDF*). This real-world analysis shows how WvD's methodology produces CI rankings that are not generally robust to alternative scalings of the SRHS data. Additional results presented in section 4 illustrate how the WvD technique may be inappropriate when comparing qualitative health data generated by survey questions that are not identical, or when comparing qualitative health data from different cultures -- two common circumstances in international comparisons. Section 5 provides a summary of findings, discusses the general limitations of SRHS data, and points ahead to the qualitative health measures that will be introduced in chapters III and IV.

2. Interpreting the health concentration index when qualitative health data are used

With the CI, WvD intend to measure inequalities in health that are "systematically related to socioeconomic status,"[4] so as to determine "whether socioeconomic inequalities in health are more pronounced in some countries than in others."[5] Standard economic theory for the demand of health predicts that health will increase with income, education, wealth, and other measures of social status. Presumably, WvD do not intend their method to be used to test this theory, i.e., to measure the extent to which health increases monotonically with socioeconomic status (SES). Rather, WvD presumably intend the CI to measure the *degree* to which health

[4]Wagstaff et al (1991), p. 546.

[5]Van Doorslaer et al (1997), p. 94.

increases with SES, i.e., the severity or intensity of socioeconomic inequality. One way to characterize the severity of socioeconomic inequalities in health would be to calculate the percentage of total inequality attributable to SES by decomposing some overall measure of health inequality into within- and between-socioeconomic group (SEG) inequality. Yet Wagstaff, Paci, and van Doorslaer (1991) criticize the decompositional approach for its inability to attribute between-SEG inequality to a clear, positively-sloped, socioeconomic class gradient.

Instead, WvD develop and use a concentration index for health and income that is an adaptation of Kakwani's (1977) tax progressivity index. Kakwani's index measures the area between the standard Lorenz curve for pre-tax income and the tax concentration curve, which plots the cumulative proportion of the population, ranked by pre-tax income, against the cumulative proportion of total tax payments. If tax payments are strictly proportional to income, then the tax concentration and Lorenz curves coincide. In such a case, Kakwani's index would be equal to zero. Kakwani's index of progressivity, P^K, is defined as: $P^K=CI^{tax} - G^{inc}$, where G^{inc} is the Gini coefficient and CI^{tax} is the Gini-like tax CI, equal to twice the area between the diagonal and the tax concentration curve in a Lorenz diagram.

The health concentration curve is analogous to the tax concentration curve, except that the health concentration curve plots the cumulative proportion of health (or ill-health, if desired), rather than the cumulative proportion of taxes paid, against the cumulative proportion of the population ranked by income (i.e., relative income rank). As Figure 1 illustrates, if health is equally distributed across the population, the concentration curve is equal to the diagonal. Consequently, the index associated with the

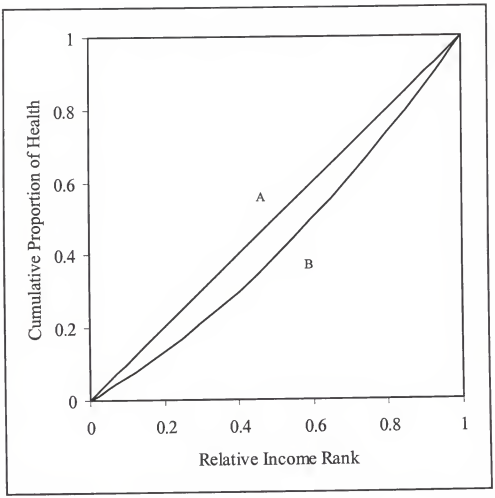


Figure 1. The health concentration curve for two hypothetical populations. In population A, health is distributed equally across income groups and the concentration curve is equal to the diagonal. In population B, richer groups have a greater proportion of overall health. The health CI is equal to twice the area between concentration curves A (the diagonal) and B.

health concentration curve, CI^{health}, is equal to twice the area between the health concentration curve and the diagonal.

2.1 Calculating the health CI

The health CI (the superscript is heretofore dropped) is designed to measure relative differences in health across the socioeconomic strata of a population. As with the Gini coefficient, the CI has a formula for continuous distributions and a formula for discrete distributions. Many empirical applications use categorical data and lend themselves to the discrete form

$$CI = \frac{2}{\mu_h} \sum_{t=1}^{T} g_t \mu_{ht} R_t - 1 \tag{1}$$

where g_t is the fraction of the population in socioeconomic group (SEG) t, μ_h is the mean level of health in the population, μ_{ht} is the mean level of health in SEG t, and R_t is the relative (or percentile) rank of SEG t, calculated as

$$R_t = \sum_{\gamma=1}^{t-1} g_\gamma + \frac{1}{2} g_t \tag{2}$$

The CI is calculated using health data in the form of SEG and population means, μ_h and μ_{ht}. If income data are not grouped, then T=N, where N is the size of the population, and R_t is simply the proportional rank of person t.[6] As described above, SEGs are ordered

[6]Kakwani, Wagstaff, and Van Doorslaer (1997) note that the value of the CI is dependent upon the number of SEG groups T, i.e., when the income ordering is incomplete. Clearly, Var(R) will differ across surveys with different categorizations of income, introducing potential error in CI rankings. This type of error is ignored in the present analysis since it has been previously noted by the principle authors of the CI.

from lowest to highest status. The CI can be calculated for health (where larger positive numbers indicate greater socioeconomic inequality), or ill-health (where larger negative numbers indicate greater socioeconomic inequality). WvD measure the CI for ill-health.

2.2 Dependence of the CI on the distribution of health

Interpreting the CI requires some assumption regarding the nature and quality of the data used to compute it. Ideally, the CI would be calculated using a continuous measure of individual health status that has a unique numeric scale.[7] This section interprets the CI assuming that such ideal health data are available, thereby casting the CI in its best light and abstracting from the data-specific concerns addressed in later sections. The discussion highlights the dependence of the CI on key attributes of the distribution of health.

According to Wagstaff et al (1991), the health CI when T=N can be interpreted as

$$CI = 2Cov(R,h)/\mu_h \tag{3}$$

where R is the relative income rank of each individual and μ_h is the mean health level for the population. Wagstaff et al (1991, 1994) and Kakwani et al (1997) show that the CI can be expressed partly in terms of β, the OLS coefficient on R in a regression explaining variations in health

$$h_i = \alpha + \beta R_i + \epsilon_i \tag{4}$$

---

[7]It is not clear that a 'natural' numeric scale representing general health status exists. Nor is it clear that a unique and meaningful numeric scale can be devised. Notice, for example, the continuing lack of agreement in the cost-effectiveness literature as to the appropriate weights to assign to a healthy year of life when calculating Disability (or Quality) Adjusted Life Years.

for individual data, where i = i...N, or

$$\sqrt{n_i}\,h_i = \sqrt{n_i}\,\alpha + \sqrt{n_i}\,\beta R_i + \varepsilon_i \tag{4a}$$

for data grouped into t = 1...T income categories. Since

$$\beta = Cov(R,h) \,/\, Var(R), \text{ then} \tag{5}$$

$$CI = 2Var(R)\beta / \mu_h \tag{6}$$

from (3) and (5) above. Note first that with an approximately complete income ranking (i.e., T approaches N) and a large sample, Var(R) approaches 1/12, the variance of a uniformly distributed continuous random variable defined over the range [0,1]. Under these conditions, the CI is directly proportional to the coefficient β, normalized by the population's mean level of health, μ_h

$$CI = \beta / 6\mu_h \tag{7}$$

Consequently, when h is an unconstrained continuous measure of individual health and T is very large, the CI carries with it a graphical interpretation as the degree to which health is disproportionately concentrated among those with relatively higher incomes, and a statistical interpretation as (a proportion of) the relative increase in health due to an increase in income *rank*. The statistical interpretation emphasizes the relationship between the CI and β / μ_h, which is also known as the Relative Index of Inequality (RII).[8]

Interpreting of the CI in this ideal empirical setting can be aided by substituting for β the correlation between income rank and health, ρ_{Rh}, where

$$\rho_{Rh} = \beta * SD_R / SD_h \cong .289 * \beta / SD_h \,, \tag{8}$$

---

[8]See Wagstaff, Paci, and van Doorslaer (1991).

since the standard deviation of income rank, SD_R, approaches $\sqrt{1/12} \cong .289$. Substituting (8) into (7) yields

$$CI = 2 * \sqrt{1/12}\, \rho_{Rh} SD_h / \mu_h \cong .577\, \rho_{Rh} SD_h / \mu_h . \tag{9}$$

Note that SD/μ_h is the coefficient of variation in health, CoV_h, a measure of relative inequality in the distribution of health, so that

$$CI = .577 \rho_{Rh} CoV_h . \tag{10}$$

Before moving on to discuss the CI in a more realistic empirical framework, it is worth pausing to evaluate the policy implications of using the CI when approximately continuous and numerically scaled health data are available. It is apparent from (10) that socioeconomic inequality in health as measured by the CI could differ in two populations because of: (a) differences in relative health inequality in the two populations (CoV_h); (b) differences in the proportion of the variance in health explained by income rank (ρ_{Rh}); or (c) differences in both relative health inequality and the proportion of variation in health explained by income rank. Thus, even when continuous and numerically scaled health data are used, it may be difficult to interpret and act upon results of CI comparisons without also looking more closely at the attributes of the distribution of health.

The potential ambiguity in CI rankings is illustrated below with an example. Consider two hypothetical countries A and B, each with a population of one thousand and a mean level of health of 7.5. In both countries a one percentage point increase in income rank provides a health boost of .05, so that β, the coefficient on income rank in (4) is 5. According to the definition in (7), the CIs of the two countries must both be

equal to 0.11, suggesting that socioeconomic inequalities are equally pronounced in each country. Nevertheless, it is possible that the bivariate distributions of health and income rank for A and B could differ substantially. The data for countries A and B illustrated in Figure 2 were obtained by computing (4) using randomly-generated error values. Consistent with the present example, it is assumed that n=1,000, $E(h|x)=7.5$, $\beta=5$, and that the error term is distributed normally with a mean of zero and a standard deviation of 1.5 for country A and .5 for country B. As Figure 2 illustrates, the level of health inequality is much higher in country A, while the correlation between income rank and health is clearly stronger in country B.[9] These underlying differences in the distribution of health and the health-income relationship completely offset each other, since the CI is approximately the same in country A and country B.[10] The example highlights the weakness of the CI as a policy tool, for even though the CI fails to distinguish between the two countries, it is difficult to imagine that policy makers would be indifferent as to whether country A or country B was more deserving of intervention.

Equations (9) and (10) emphasize the dependence of the CI on measured characteristics of the distribution of health, raising questions about how qualitative health data might be used to compute the CI, and how the resulting CI values and associated rankings might be interpreted. The next two sections address these questions, first by describing the way that WvD use self-reported health status data to compute the CI, and then by evaluating the CI in this more realistic empirical framework.

---

[9]The CoV_h for country A is .28, compared to .20 for country B. The correlation between income rank and health is .69 in country A and .94 in country B.

[10]The computed CIs are .1092 for country A and .1091 for country A.

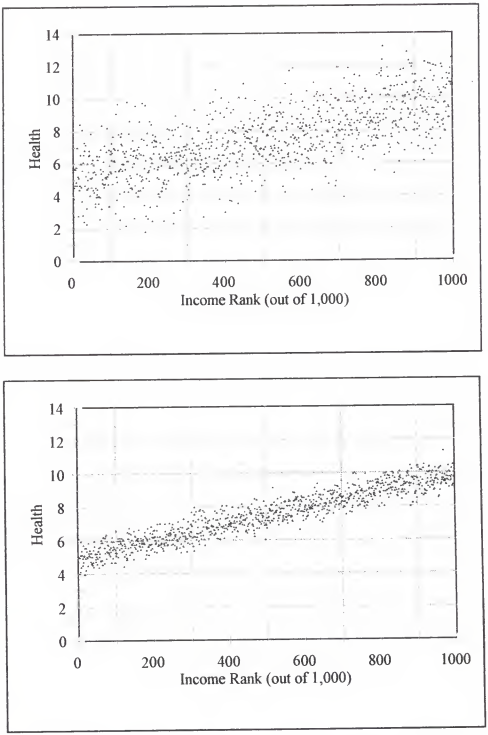


Figure 2. Plot of health against income rank for two randomly-generated distributions. Country A's distribution is in the upper panel, country B's in the lower panel.

3. Interpreting the CI when qualitative health data are used

While other morbidity or mortality data have also been used to calculate the CI, the primary data WvD use for this purpose is SRHS.[11] A key challenge in using SRHS to calculate the CI is the need to map the categorical data onto a numeric scale so that SEG and population mean health levels can be calculated. The principle alternative to the WvD methodology is to calculate socioeconomic inequality using SRHS data that have been dichotomized into observations that are in ill health and those that are not. For example, Kunst et al (1995) calculate the RII (see section 2.2) using a continuous variable representing the odds that a person in socioeconomic group t≠T will have less than "good" SRHS, expressed as a ratio of the odds that a person in socioeconomic group T will have less than "good" SRHS. WvD appropriately criticize this method because the inequality rankings it generates are not robust to the analyst's choice of a threshold separating 'good' from 'ill' health. Moreover, they note that dichotomizing reduces the distributional information imparted by SRHS data, "resulting in a loss of cross-sample variation" since the number of possible health values is reduced from four or five down to two.[12]

3.1 The WvD methodology for scaling SRHS data

The WvD approach to scaling the SRHS data is to assume that SRHS responses

[11]See for example Wagstaff and van Doorslaer (1998), who calculate CIs using SRHS as well as several other health indicators, and Wagstaff et al (1991), who calculate CIs using morbidity as well as mortality data.

[12]Wagstaff and van Doorslaer (1994), p. 281.

are generated from an underlying continuous latent variable, h*. It is assumed that h* is distributed as a standard lognormal variable in each of the samples for which the CI is calculated. Wagstaff and Van Doorslaer (1994) provide at least two motivations for assigning each sample the same distribution of latent health:[13]

- other measures of ill-health, e.g., morbidity, are observed to be skewed, with few reporting ill-health and many reporting good health; and
- since WvD wish to compare SRHS data from several surveys, but the number of response categories J varies across surveys, the assignment of scale values to categorical responses should be independent of the number of categories.

The basic approach WvD use to obtain health scores consistent with the standard lognormal distribution is to calculate the inverse cumulative distribution function for h* at each individual's relative health rank. Since SRHS data are highly aggregated, WvD partition h* into J ranges and assign each individual in the j^{th} category a representative value for h* in the j^{th} range of the distribution. The representative value they use is the geometric mean. It is unclear why WvD choose the geometric mean as the representative

[13]Wagstaff and van Doorslaer (1994) also motivate the distributional restriction by noting that it standardizes each population to a common mean level of latent health, which is desirable since WvD's research interest is to compare *relative* socioeconomic inequality in different populations. Nevertheless, as equations (1), (3), (7) and (9) illustrate, the CI itself measures relative differences in health across socioeconomic strata, so that assigning each population a common mean level of health has no effect on the CI.

value for each range of the distribution, except as a way to overcome the difficulty in evaluating the integral function for the arithmetic means of these ranges.[14]

3.2 Computing scale values for SRHS data

WvD assume the existence of an underlying continuous variable h* that generates SRHS responses in J categories as follows

$$\begin{aligned} &\text{SRHS} = 1 \text{ if } -\infty < h^* \leq \alpha_1 \\ &\text{SRHS} = 2 \text{ if } \alpha_1 < h^* \leq \alpha_2 \\ &\vdots \\ &\text{SRHS} = J \text{ if } \alpha_{J-1} < h^* \leq +\infty \end{aligned} \tag{11}$$

where α_j is the threshold separating category j and j + 1. The representative values associated with each SRHS category j are identified by first dividing the area under the distribution h* into J ranges according to the proportion of the population in each SRHS category. The geometric mean of each of these ranges, h_j^*, is designated to be the representative level of health for each SRHS category j. To compute the geometric means, WvD first use the SRHS data in the manner described above to partition the area underneath a standard normal distribution, increasing in good health. WvD then use the arithmetic mean of each range underneath the standard normal distribution to calculate geometric means for each categorical range, an inconsistency that is addressed below in section 3.3.

The threshold levels of latent health that constitute the upper bounds of the j^{th}

---

[14]The exact conditional arithmetic means of a standard lognormal distribution cannot be obtained. Such means must be calculated numerically.

categorical range of the standard normal distribution are

$$\hat{\alpha}_j = \Phi^{-1}\left(F_j\right), \text{j=1,...,J, where} \tag{12}$$

$$F_j = \sum_{i=1}^{j} \frac{n_i}{N}, \tag{13}$$

Φ is the cumulative distribution function for the standard normal distribution, n_i is the number of observations in SRHS category i, and N is the total number of observations in the sample. In equation (13), F_j is the cumulative proportion of the population in group j or less and can thus be thought of as the health rank for those in SRHS category j.[15] Since health in this initial step of the process is distributed standard normal, the arithmetic mean health score for each SRHS category j in a population is

$$\hat{z}_j = \left(\frac{N}{n_j}\right)\left[\phi(\hat{a}_{j-1}) - \phi(\hat{a}_j)\right] \tag{14}$$

where ϕ is the standard normal density.

To calculate the geometric mean of range j of a standard *lognormal* distribution, one simply exponentiates the mean health scores $\hat{z}_j$ obtained in (14). Since the lognormal distribution is skewed to the right when health is increasing, and since health

[15]The reader may notice that this definition of health rank does not mirror the definition of income rank, R_t in (2). Whereas R_t represents the rank of the representative, i.e., middle, person in SEG t, F_j denotes the proportion of the population in SRHS group j or less. Since it is only necessary to identify each individual's health score, and each individual in SRHS j has the same level of health, it is not necessary to identify the middle person in SRHS group j.

is thought to be skewed in the opposite direction (with few in poor health), the sign of each $\hat{z}_j$ score is reversed to obtain the representative ill-health scores for each SRHS category

$$h_j^* = e^{-\hat{z}_j}, \quad j=1,...,J, \tag{15}$$

which are the geometric means of each categorical range of the standard lognormal distribution.

3.3 Computing the CI using scaled SRHS data

WvD calculate the mean health of SEG t by applying the group-specific SRHS distribution to the representative values calculated in (15), so that

$$\mu_{ht} = \sum_{j=1}^{J} f_{tj} h_j^*, \tag{16}$$

where f_{tj} is the proportion of group t in SRHS category j. For the population as a whole, WvD calculate mean health as follows

$$\mu_h = \sum_{j=1}^{J} f_j h_j^*, \tag{17}$$

where f_j is the proportion of the population in SRHS category j. The SEG and population means are then used to calculate the CI, as in (1). Note, however, that the arithmetic SEG and population means in (16) and (17) are functions of the conditional geometric means calculated in (15). Since geometric means do not aggregate up to arithmetic means, the population mean in (17) will not generally be equal to the imposed mean of the standard lognormal distribution, which is about 1.649. For example, the computed

population mean level of SRHS in the USA in 1994, using data from the *NHIS*, is 1.58 according to the WvD methodology.[16] This inconsistency in the WvD methodology could be alleviated either by numerically calculating the arithmetic means of each categorical range of the standard lognormal distribution, or by using a different distribution that facilitates exact computation of conditional means. For example, the analysis in section 4 below assumes that latent health follows the gamma distribution, for which exact conditional means can be computed.

3.4 Interpreting the CI in the context of equivalent health distributions

To evaluate the WvD approach to the use of categorical health data, it is helpful to consider first the method's attributes when applied to data that yield a complete health ordering, i.e., that yield a continuous distribution with indeterminate parameters.[17] This assumption casts the CI in the best possible light *given* that each population has been assigned the same distribution of health, as dictated by WvD's scaling methodology.

If a complete health ordering is available, and if continuous income data are also available, then J=T=N. In this case, the process of converting categorical data into scaled values described in section 3.2 need not be followed, and the inverse CDF of the

[16]The NHIS data are described in detail in chapter III, sections 4 and 5. Also note that in every state included in the examples in section 4 of this chapter, the computed means were less than 1.649.

[17]A complete health ordering might be obtained, for example, by using predicted values of latent health as generated by a principal components analysis (PCA) or a multiple indicator, multiple cause (MIMIC) model. Latent measures of health in such models may produce a nearly complete ordering of the sample population, though the numeric values of health cannot be said to have cardinal significance.

standard lognormal distribution can be calculated directly at each individual health rank with minimal error. Direct calculation of latent health scores in this manner prevents the inconsistency identified in section 3.3 that arises when highly aggregated health data are used to calculate SEG and population means. Let the standard lognormal distribution be represented by Θ. Then individual latent health scores are completely determined by health rank F_t and by Θ as follows

$$h_t = \Theta^{-1}(F_t). \tag{18}$$

Wagstaff and van Doorslaer (1998) briefly discuss the potential sensitivity of CI rankings to the choice of distributional form. Nevertheless, the profound effect that this choice can have on the interpretation of the CI has not been acknowledged in the literature. Note that since the assigned standard lognormal health distribution has a constant variance of $\mathrm{Var}(h_t) = \mathrm{Var}(\Theta^{-1}(F_t)) = 4.671$ and a constant mean of $\mu_h = 1.649$, and since the standard deviation of income rank is approximately equal to 1/12, the correlation between income rank and health (8) may be re-written as

$$\rho_{Rh} \cong .289\beta / SD_h \cong .289\beta / 2.161 \cong .134\beta \tag{19}$$

which reduces the concentration index (9) to

$$CI \cong .577\,\rho_{Rh} SD_h / \mu_h \cong .757\,\rho_{Rh} \tag{20}$$

Substituting $\Theta^{-1}(F_t)$ for h_t yields

$$CI \cong .757\rho_{R\Theta^{-1}(F)}. \tag{21}$$

Equation (21) demonstrates that imposing a common distribution of health on each compared population results in a CI that is a fixed proportion of the correlation between income rank, R, and the assigned scale values, $\Theta^{-1}(F_t)$.

Of course, CI values might change if a distribution *other* than Θ were used in the WvD scaling process, or if some other scaling methodology were employed (these inferences will be demonstrated in section 4.1 using SRHS data). Note that when the function Θ returns the health rank itself, the correlation in equation (21) is equal to the Spearman Rank Correlation for health and income.[18] Indeed, the only data WvD use to calculate the CI are the health and income rands for a population. Consequently, unless Θ is believed to be the distribution of health in all populations for which the CI is calculated, then CI rankings generated by the WvD method may be described as a scale-(or distribution-) dependent rank correlation between income and health, i.e., *a scale-dependent measure of how uniformly health improves with income.* WvD's success in going beyond such an interpretation to measure, as intended, the degree to which health increases with income rank, is thus a function of how well Θ approximates the distribution of health in each population.

4. Evaluating the CI using real-world data

The previous sections have considered the limitations of the CI generally and in the context of continuous but un-scaled health data. Nevertheless, WvD obtain the CI from partial health orderings using highly aggregated SRHS data. This raises questions about the nature of CI rankings in a real-world empirical setting. In particular, how does the WvD scaling methodology, described in section 3.2, affect CI rankings when there are only four or five SRHS categories, as is normally the case? Also, how sensitive are

[18]Woolson (1987).

CI rankings to the choice of distribution (or scale) to be imposed on each population's health profile? Would rankings differ if a direct scaling method were used? Finally, how comparable are SRHS data across countries or surveys, and how might any incomparability affect CI rankings?

Using data from the 1994 National Health Interview Survey State Data Files (*NHIS-SDF*),[19] this section provides examples that demonstrate three types of non-robustness in CI rankings generated by the WvD method. The first example demonstrates that CI rankings are sensitive to the choice of distribution (or scale) for the SRHS data. The second demonstrates that CI rankings are sensitive to the number of SRHS categories in the survey data used. The third shows that CI rankings are also sensitive to alternative labeling of the SRHS categories.

4.1 Sensitivity to the choice of scale or distributional form

To illustrate the sensitivity of CI rankings to alternative scalings of the SRHS data, *1994 NHIS-SDF* data for five selected states and the nation as a whole are used to calculate CIs according to the methodology described in sections 2 and 3 above. Following WvD, each socioeconomic group t in the example below is given the age and sex composition of the state as a whole in order to account for co-variation between age, gender and income.[20] As a point of reference, the CI was first calculated using WvD's

[19]A description of this data is included in sections 4 and 5 of chapter III.

[20]Data from this survey is highly aggregated to preserve confidentiality: income is reported in four categories, while ages are reported in six and gender in two. Hence, 6*2=12 demographic categories, d=1,...,12, form the basis for the standardization. The standardized proportion of group t comprised of individuals of demographic type d is

mean-inconsistent methodology (see section 2.3). Results of these CI calculations are displayed in column 1 of Table 1. New York has the lowest CI, in absolute value, and hence the least amount of socioeconomic inequality. New York is followed by California, Ohio, the USA as a whole, Texas, and Tennessee, which has the highest CI and hence the greatest amount of socioeconomic inequality. Next, the CI was calculated using geometric means for SEGs and for the population as a whole, eliminating the mean-inconsistency. The results are shown in column 2 of Table 1. Using this mean-consistent approach, the rank ordering of California and Ohio is reversed. Nevertheless, there may be little justification for using the geometric mean to compute the CI. An alternative method for generating mean-consistent CIs is to use a different functional form for the assigned distribution of h*, such as h*~ gamma(2,1), in place of the standard lognormal distribution.[21] The use of the gamma distribution enables the exact computation of arithmetic means for categorical ranges, which can then be used to compute arithmetic SEG and population means as in (16) and (17). The resulting CIs are

---

equal to the proportion of type d in the general population, n_d/n. The income group means (16) used to calculate the CI (1) were obtained using $\mu_{ht} = 1/n_t \sum_{j=1}^{5} h_j^* \sum_{d=1}^{12} n_{tdj} \left(\frac{n_d}{n} \Big/ \frac{n_{td}}{n_t} \right)$, where n_t is the number of people in income group t, h_j^* is the latent health score for those in SRHS category j from (15) above, n_{tdj} is the unstandardized number of individuals in group d of income group t that have a latent health score of h_j^*, and n_{td}/n_t is the unstandardized proportion of income group t that is in demographic group d. Mean health for the country as a whole was calculated by using (17), where the proportion of the population with latent health h_j^* is calculated using the standardized totals $f_j = \sum_{t=1}^{4} \sum_{d=1}^{12} n_{tdj} \left(\frac{n_d}{n} \Big/ \frac{n_{td}}{n_t} \right)$.

[21] This distribution is somewhat less skewed than the standard lognormal distribution.

Table 1. CI rankings under different scalings of SRHS. The SRHS scales used are increasing in ill-health except IV.

	I WvD (log-linear; arith/geo)		II WvD (log-linear; geo/geo)		III WvD (gamma/concave)		IV WvD (gamma/convex)		V Direct (convex)[1]	
	CI	CI Rank	CI	CI Rank	CI	CI Rank	CI	CI Rank	CI	CI Rank
CA	(0.1412)	2	(0.1034)	3	(0.0925)	3	0.0832	5	(0.0695)	4
NY	(0.1289)	1	(0.0963)	1	(0.0816)	1	0.0704	1	(0.0619)	1
OH	(0.1511)	3	(0.1019)	2	(0.0907)	2	0.0707	2	(0.0645)	2
TN	(0.1655)	6	(0.1151)	6	(0.1056)	6	0.0921	6	(0.0725)	5
TX	(0.1595)	5	(0.1093)	5	(0.0991)	5	0.0805	4	(0.0727)	6
USA	(0.1528)	4	(0.1066)	4	(0.0952)	4	0.0791	3	(0.0692)	3

Notes

1. This scale, [5, 4.51, 3.77, 2.66, 1], has the property that each intercategory gap is two-thirds as large as the next gap.

shown in column 3 of table 1. Again, the ranking of California and Ohio are reversed as compared to the ranking using the mean-inconsistent WvD approach.

The remaining two columns of Table 1 illustrate the effects of using alternative distributions – and, hence, different numeric scales for SRHS data. Column 4 shows the CI rankings assuming that h* is distributed ~gamma(2,.5) in *good* health. Unlike the (implicit) scalings used in the CI rankings in columns 2 and 3, the implicit scale created by assigning the gamma(2,.5) distribution to each state's population assumption emphasizes distances between the upper health categories. Though there may be little justification for the use of a distribution skewed in this manner, the results are presented here as an illustration of the potential sensitivity of CI rankings to the choice of distribution. Using the gamma(2,.5) distribution to compute the CI alters the rankings substantially, dropping California's rank by two and raising the rankings for Texas and the USA as a whole. An alternative scaling methodology would be to assign scale values to SRHS categories directly, without assuming the existence of an underlying latent (and continuous) distribution of health. Note that this approach to scaling SRHS data differs from the WvD methodology in that SRHS categories are assigned a scale value that holds across all populations. An advantage of this approach is that it allows for variations in mean health and dispersion in health across populations.[22] An example of this approach is presented in column V, for which the scale [5, 4.51, 3.77, 2.66, 1] = [Poor, Fair, Good Very Good, Excellent] has been used. Like the implicit scale used in column IV, this direct scale emphasizes differences between the upper health categories. Nevertheless,

[22]Allowing the mean to vary has no effect on the CI, a relative measure of inequality.

the CI rankings have been altered once again, with re-rankings observed among California, Texas, and Tennessee, as compared to the rankings in column IV.

Taken together, the results illustrated in Table 1 demonstrate that the CI is not generally robust to alternative distributional assumptions, as when using the WvD scaling methodology, nor to the use of alternative scaling methodologies.[23]

4.2 Sensitivity to the number of SRHS categories

Data from the same five states and the USA are also used to illustrate the sensitivity of CI rankings to the number of SRHS categories J when the WvD scaling approach is employed. To demonstrate a lack of robustness to the number of categories, observations in adjacent categories in three states are combined, and the latent health scores, h_j^*, group means, μ_{ht}, population mean, μ_h, and CIs are recalculated using J-1 categories of health. In each case, it is assumed that h*~gamma(2,1), so that arithmetic means may be used in all computations. If the CI values are different when calculated with J and J-1 categories, it suggests that two equivalent populations can have different CIs if calculated using two different samples with, respectively, J and J-1 SRHS categories.

The results indicate that in every case where categories are combined, including those not presented, the CI values are different after combining adjacent categories. If the CI values change enough when the two adjacent categories are combined, then the

[23]The results shown in Table 1, along with additional results not shown, suggest CI rankings among the selected states may be consistent when scales emphasizing differences among the lesser SRHS categories are used (i.e., concave scales). Nevertheless, it is not apparent how one might demonstrate such robustness.

state's ranking versus other states with J categories can be altered. Such re-rankings obtain in three circumstances, as shown in Table 2. In the first, the excellent and very good categories are combined and the CI for California is recalculated. The result is that the (absolute) value of the CI for California falls from -.0925 to -.0864, and its health inequality ranking rises from 3rd to 2nd. When the very good and good categories are combined for Ohio, its CI falls from -.0907 to -.0793, and its ranking rises from 2nd to 1st. Finally, when the good and fair categories are combined for Texas, its CI falls from -.0991 to -.0909, and its ranking rises from 5th to 3rd. Apparently, the CI is not robust to the number of SRHS categories used in the sample.[24]

4.3 Sensitivity to SRHS category labels

The dependence of CI rankings on the number of SRHS categories limits the applicability of the approach in international settings. There is, however, an even more fundamental problem associated with the use of qualitative data to generate international comparisons. Recall that according to the WvD scaling method, the thresholds α_j separating categories j and j + 1 in the underlying distribution of health are direct functions of the cumulative distribution of SRHS. Assume for the moment that surveys

[24]Kakwani et al (1997) find the CI to be lower when computed using a dichotomous chronic illness indicator than it is when 4-category SRHS data are used. They interpret this finding by suggesting that using a dichotomous chronic illness indicator results in a loss of information in the severity of illness, thereby reducing measures of health inequality. My findings suggest that in most cases, a reduction in the number of categories of health will reduce the value of CI. One possible explanation for this is that since there is a positive relationship between health and socioeconomic status, any smoothing in the concavity of the implicit WvD scaling (the typical result when fewer categories are used), is predicted to bring health means of SEGs closer together.

Table 2. CI ranks with various SRHS categories combined.

	5 SRHS categories		4 SRHS categories		
	CI	CI Rank	Categories combined	CI	CI Rank
USA[1]	(0.1528)	4th	Very Good and Good	(0.1426)	3rd
OH	(0.1511)	3rd	Good and Fair	(0.1344)	2nd
TN	(0.1655)	6th	Fair and Poor	(0.1373)	2nd

Notes:

1. van Doorslaer et al (1997) find the USA to have a CI of (0.1360) using data from the 1987 *National Medical Expenditure Survey* where the four categories are "Excellent, Good, Fair, and Poor" as in this example.

for each country to be compared by the CI have an SRHS question with J categories. Given that WvD assign each country's population the same distribution of health, the thresholds separating categories can only differ across countries if respondent conceptions of the meaning of each category differ. The most obvious reason that category meaning might differ across surveys is labeling, e.g., if the highest category is labeled "excellent" in one country's survey, "very good" in another's, and "good" in yet another's.[25] In each of these cases, the location of the threshold between the highest and next highest categories is likely to be different when measured in the dimension of the underlying latent health variable; the threshold is likely to be highest when the highest category is labeled excellent.

Indeed, if the same distribution of health is imposed on all populations, the location of between-category thresholds is *solely* a function of category labels or their interpretation by different cultures. This sort of categorical non-alignment leads to a fundamental inconsistency in distributional comparisons involving samples from different countries. If, for example, the percentage of a sample that is in the highest category differs across two countries, A and B, then the threshold α_{J-1} separating the two highest categories will be different in the two countries. For health ranks between α_{J-1}^{A} and α_{J-1}^{B}, the health score h_J^* assigned to individuals will depend on which country they are in: individuals in one country will be assigned h_J^*, the mean health score for the highest category, while individuals in the other country will be assigned h_{J-1}^*, the mean

[25]Surveys included in Van Doorslaer et al's (1997) international comparison of CIs include just such a range of categorical labeling, translated as they are from the original languages.

health score for the next highest category. Hence, individuals who are presumed to have the same (relative) level of health are assigned different (relative) levels of health because the individuals responded differently to the (potentially different) labeling of SRHS categories. Apparently, two countries with exactly the same underlying distribution of latent health can have different observed distributions of SRHS due to categorical labels or their interpretation by respondents.

The following example illustrates that the dependence of individual health scores on arbitrary differences in categorical labeling can alter CI rankings. Two different sets of adjacent categories are combined for a sample of data from the state of New York, in each case reducing the number of SRHS categories from five to four.[26] Combining the excellent and very good categories results in a CI of -.0733. Combining the fair and poor categories results in a CI of -.0799. This suggests that the CIs for two identical countries (or states) could differ even when four-category SRHS data are collected in each country, if the labels on the categories are different or if respondents interpret the meaning of the category labels differently. In the case of New York, if the population is asked whether their general health status is poor, fair, good, or "better than good," the resulting CI value will likely be different than if the same population is asked whether their health is "less than good," good, very good, or excellent.

Health scores in the WvD methodology are thus seen to be functions of categorical labels, which may differ arbitrarily across countries and surveys.

---

[26]As in section 4.2, the presumed distributional form in this example is h*~gamma(2,1).

4.4 Discussion

WvD propose their method of using scaled SRHS data to calculate the CI as a way to alleviate the sensitivity of socioeconomic inequality rankings to the subjective choice of a break point in a dichotomized SRHS variable, and as a means of facilitating international comparisons when survey questions differ. WvD claim that their transformation of SRHS data using the standard lognormal distribution yields a scale that is independent of the number of SRHS categories.[27] This is true, however, only in the sense that the (assigned) underlying *continuous* health score is not affected by the number or labeling of SRHS categories. It is apparent that partitioning the underlying health distribution into a small number of SRHS categories does not affect the location of the observations in that underlying dimension. Such partitioning only affects the category each observation is assigned to and the mean level of health in that category, i.e., the scale value assigned to each SRHS category. As a consequence, however, the assigned health scores used to calculate the CI are not independent of the number of SRHS categories, nor their labels. As demonstrated by the examples above, inequality orderings using the CI, like those generated by the Gini coefficient, the Robinhood index, or other inequality indices, are not necessarily robust to these numeric assignments,[28] so that WvD's results are potentially invalid where the number of categories, their labels, or respondent interpretation of these labels differ from one population to the next.

---

[27]Wagstaff and van Doorslaer (1994), p. 284.

[28]As a rule, inequality indices are either 'absolute,' i.e., rankings are independent of additive transformations of the data, or 'relative,' i.e., rankings are independent of multiplicative transformations of the data, but not both.

Note that these arbitrary variations in CI rankings could potentially be avoided if it were possible to align survey questions so that labels imparted a similar meaning to each sampled population. Even if they were so aligned, the examples in section 4.2 demonstrate that CI rankings are also contingent upon the particular distribution or scale assigned to the SRHS data. The scale dependence of CI rankings may be unimportant if WvD have chosen the 'correct' scale -- or the correct scaling methodology -- for SRHS data. WvD's methodology assumes that general ill-health is distributed according to the standard lognormal distribution in every population. The only justification WvD offer for this assumption is that the distributions of other qualitative health indicators appear to be highly skewed. It is difficult to see, though, how one set of qualitative health data could serve as a numeric benchmark for another. Moreover, WvD provide no evidence supporting their assumption that ill-health is distributed equivalently in every population.

Might WvD, despite these concerns, have selected the 'correct' distribution for SRHS data? For a Danish sample used in Wagstaff and van Doorslaer (1994) the mean level of health assigned to those in the excellent category of health according to WvD's methodology is .28, while the mean health score for those in poor health is 10.52: excellent health is assigned a value 1/37th the value assigned to poor health. Also, the distance between poor and fair health in this derived scale is more than 6 times the distance between very good and excellent health. Gerdtham et al's (1998) comparison of numeric and categorical self-reports in a Swedish sample demonstrates that at least one sample of respondents might disagree with the extreme concavity in WvD's implicit scales. For this Swedish sample, the WvD method yields an even more highly concave scale (in ill-health): the ratio of excellent health is assigned a value 1/42nd the value of

poor health, and the difference between poor and fair health is 14.4 times the difference between very good and excellent health. The comparable ratios for numeric self-reports (using the direct rating scale method) are 1/11th and 4.4, a far less concave scale.[29]

5. Conclusion

This chapter offers a broad review of WvD's use of qualitative SRHS data to calculate the CI. Several concerns with the WvD approach are identified. First, the chapter demonstrates that unless WvD are correct in assigning the same distribution of health to each population, the resulting CI may be described as a scale-dependent measure of how uniformly health improves with socioeconomic status, e.g., income. Second, it is shown that WvD's method for using highly aggregated SRHS data to make international comparisons can result in arbitrary variations in CI rankings. Specifically, it is shown that CI rankings generated using the WvD scaling methodology are sensitive to the number of SRHS categories included in the survey and to the labeling of each category, two characteristics which often differ from one survey to the next. Finally, the chapter demonstrates how WvD's use of the standard lognormal distribution results in an inconsistent aggregation of group means -- an inconsistency that could be alleviated with

[29]Despite the apparent inconsistencies between numeric SRHS data and WvD scalings of categorical SRHS data for a Swedish sample, Gerdtham et al find no statistical differences between the CI scores computed using the two numeric health scales. The authors suggest that their findings validate the WvD methodology of imposing a standard lognormal form on the distribution of health. This validation of the WvD methodology, however, would appear to be insufficiently generalizable. Gerdtham et al's results are contingent upon the sample used. To validate the WvD method would presumably require many more such tests using data from, potentially, each country and SRHS question type for which CIs are to be computed and compared.

the choice of an alternative distribution.

With the limitations of the WvD methodology identified above, the question may arise: how can CI rankings such as those presented in Van Doorslaer et al (1997) be interpreted? What information do the rankings impart? This chapter has demonstrated that unless the distribution of health is correctly assigned in each country, international CI rankings generated by the WvD method will reflect:

1. differences in the way people in different populations interpret the bounds of the SRHS categories, e.g., based on differences in labeling or in respondents' interpretations of these labels;
2. differences in the number of categories used in surveys of the different populations; and
3. differences in the socioeconomic ordering of an assigned health distribution.

The following characterization of the CI incorporates these sources of variation explicitly into the measure's formula and illustrates its dependence on the survey used to collect the data

$$CI_m = .577\rho_{R\Theta_m^{-1}(F)}CoV_{\Theta_m^{-1}(F)} \tag{22}$$

where m indexes the survey used to collect the data and $\Theta^{-1}{}_m(F)$ represents the numeric health scores of the population, which are functions of: (i) the assigned distribution of health, Θ; (ii) the survey-specific nature of the SRHS data; and (iii) the partial health

ordering of the population, as observed in SRHS data. Equation (22) illustrates how the CI may differ from one survey to the next according to the vagaries of question design and respondent interpretation, and how the CI is insensitive to any real differences in the distribution of health across countries.

Will qualitative health data support WvD's measurement goals, i.e., to compare the *intensity* of the health-SES relationship *across countries?* It is difficult to see how one could measure the intensity of the health-SES gradient without using numerically-scaled health data, if intensity is defined as the rate at which health accumulates with social rank. It is equally difficult to see how SRHS data could be made consistent across surveys and countries, overcoming potentially vast differences in language, culture, and knowledge so as to achieve equivalent interpretation of each SRHS category in each country.

Despite the limitations of SRHS data in the context of international comparisons of socioeconomic health inequality identified above, it may nevertheless be possible to glean a great deal of distributional information from such data. The remaining chapters in this thesis address the fundamental question of how to measure the distribution of SRHS without imposing (unreasonable) distributional attributes on the data. Chapter III introduces a method for making comparisons of overall health and health dispersion by relying solely on the qualitative information imparted by SRHS data. Chapter IV extends this methodology to allow distributional comparisons that apply to any *reasonable* numeric scale that might be imposed on the data. Chapter V presents a novel approach to the explanation of between-group differences in overall levels of SRHS.

CHAPTER III

MEASURING HEALTH INEQUALITY USING QUALITATIVE DATA

By R. Andrew Allison and James E. Foster[1]

1. Introduction

Many key questions in health policy require an evaluation of the inequality in the distribution of health status across individuals in a population, and how that level of dispersion might be altered through policy reforms. Objective data on individual health status, however, are often unavailable or incomplete or are difficult to interpret, especially for groups with very low mortality.[2] Consequently, several studies have turned to data on Self-Reported Health Status (SRHS) -- which provides a respondent's own perceptions of their general health condition -- in constructing an aggregate health profile.[3]

There are a number of potential advantages of using SRHS. First, it allows the individuals themselves to determine the relative importance of the various dimensions of

---

[1]Presented at the Global Health Equity Initiative workshop on measuring inequalities in health, October 27-28, 1998, St. Catherine's College, Oxford. The authors would like to thank Sudhir Anand, William Collins, and Angus Deaton for helpful comments. Foster wishes to thank the MacArthur Foundation for their support of this research through the Network on Inequality and Poverty in Broader Perspective.

[2]A second problem is that many key health variables are expressed as rates over various groups. Using data such as these, one can only measure *between-group* inequality, and the results would be conditional on the nature of the groupings.

[3]See for example: Kunst, Geurts, and van den Berg (1995) and van Doorslaer et al (1997), who examine international differences in socioeconomic inequality in SRHS; Deaton and Paxson (1998), who evaluate the distribution of SRHS over the life-cycle; and Kennedy, Kawachi, Glass and Prothrow-Stith (1998), who study socioeconomic differences in SRHS in the United States.

health, rather than assigning arbitrary weights to the components. Second, there is a great deal of evidence that SRHS is a strong predictor of objective measures of health.[4] And third, SRHS is widely collected, and is therefore available in large enough samples as to allow meaningful analysis of a broad range of questions. There are also a number of potential problems with SRHS, including the possibility of inherent biases due to the fact that it is self-reported rather than objectively determined; the relative merits of these arguments have been discussed at some length elsewhere. [5]

A critical limitation of SRHS data in evaluating the distribution of health is its qualitative nature. Respondents assess their health in various ranked categories rather than along a given cardinal scale. The US *National Health Interview Survey* (*NHIS*), for example, has five categories for respondents: poor, fair, good, very good, and excellent. Given this central feature, a key question is how to interpret the distributional information imparted by SRHS data. In particular, to what extent can standard measurement techniques used in income inequality analysis be applied in this context?

In order to make use of the standard techniques, one must first select a scale which assigns a numerical value to each category of health. A natural choice is the "linear" scale, assigning the values "1,2,3,4 and 5" to the respective categories "poor,

---

[4]See for example Singer et al (1976), Mossey and Shapiro (1982), Idler and Angel (1990), Grant et al (1995), and Appels et al (1996), who examine the validity of SRHS as a predictor of mortality. Idler and Kasl (1995) and Marmot et al (1995) relate SRHS responses to other health indicators and outcomes.

[5]See for example Manning, Newhouse, and Ware (1982), Anderson and Burkhauser (1985), Butler, Burkhauser, Mitchell and Pincus (1987), Meredith and Siu (1995), Waidmann, Bound, and Schoenbaum (1995), Strauss and Thomas (1998), and Dwyer and Mitchell (1999).

fair, good, very good, and excellent." One can make a reasonable argument in favor of such an approach – that in the absence of any additional information about the "distance" between categories, one should simply use an equidistant scale.[6] Indeed, precedent for such a scale can be found in the work of Deaton and Paxson (1997, 1998), who examine SRHS inequality over the life-cycle.[7]

The use of a specific scale along with standard methods of evaluation can be a good starting place for analyzing health inequality. However, it is clear that any given scale can be challenged, and an alternative scale may significantly alter the initial conclusions. Put differently, any results obtained using a single numeric scale may not hold up under further scrutiny. How can we be sure that a given conclusion is robust? Is there a technique for evaluating health inequality that is independent of scale?

This paper presents an "ordinal" approach to evaluating health inequality that is robust to changes in the scale. We begin in Section 2 with a critique of "mean-based" inequality measures when the underlying variable is qualitative. In Section 3 we present our median-based approach, which views inequality as a "spread" away from the median. The overall approach is embodied in a simple diagram that (analogous to the Lorenz curve) graphically indicates when inequality rises or falls. Associated with this diagram are several basic measures of spread which are helpful in signaling and explaining changes in health inequality. Sections 4 and 5 present two applications of these tools

---

[6]This is an adaptation of the reasoning behind the well-known "Borda count" in social choice theory that was applied by Sen (1976) to poverty measurement.

[7]To be precise, Deaton and Paxson (1997, 1998) use a linear "ill-health" scale in which 5 is the lowest category and 1 is the highest. In the present paper, the order is reversed.

using U.S. data. The first examines inequality in health over the life-cycle, while the second compares health inequality across states. The final section summarizes our results and suggests promising avenues for future investigation.

Before proceeding to the analysis, we would like to emphasize that the aim of this paper is to provide a method of evaluating *overall* inequality in health, without focusing on any particular cause or justification. In this respect, we are following the traditional objective of income inequality measurement (see for example Sen, 1997), which has periodically been advocated for assessing health inequality (see, for example, Tobin, 1970, Muurinen and Le Grand, 1985, or Culyer and Wagstaff, 1993). In essence, it is an attempt to describe "what is", which logically precedes analyzing the causes of inequality or judging the extent to which observed differences are inequitable.[8]

At the same time, we recognize that there are many useful studies whose main purpose is to identify the possible causes of inequality. In labor economics, "earnings equations" provide estimates of the importance of education, race, gender and other variables to one's level of earned income, and thereby help to explain observed levels of inequality in this important component of income. For health, an analogous exercise gives an assessment of the level of "socioeconomic inequalities in health" by evaluating how an individual's or a group's level of health depends on its social or economic status.[9] These types of studies provide important clues to certain facets of health inequality,

---

[8]This point is made forcefully by Le Grand and Rabin (1986), who contend "it is surely preferable first to measure the extent of the inequality in a fashion that does not prejudge the question of causality, and then to examine the possible explanations of it."

[9] See Kunst (1999).

which become more informative when placed in the broader context of overall health inequality.

2. Problems with the mean

A given population's SRHS distribution will be represented as a vector $x = (x_1,x_2,...,x_n)$, where n is the (fixed) number of categories and x_i is the number of observations in category i. A scale is a positive vector $c = [c_1,c_2, ..., c_n]$ that assigns category i the numeric value c_i, with the property that higher health categories have higher values, i.e., $c_i > c_j$ for all $i > j$. The mean of distribution x given the scale c is defined by

$$\mu(x;c) = \frac{x_1c_1 + x_2c_2 + ... + x_nc_n}{x_1 + x_2 + ... + x_n} \tag{23}$$

Traditional measures of inequality such as the variance, the Gini coefficient, Atkinson's measures, the Theil measures, and the other generalized entropy measures are all "mean-based" in that they view inequality as deviation from the mean or are normalized using the mean. In other words, all standard measures have formulas that are substantially dependent on the mean. This has important consequences for their applicability in the present context.

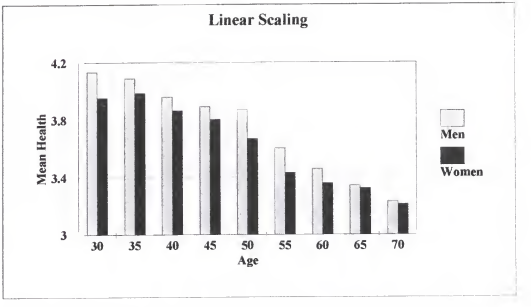
2.1 The non-robustness of the mean as an indicator of societal health

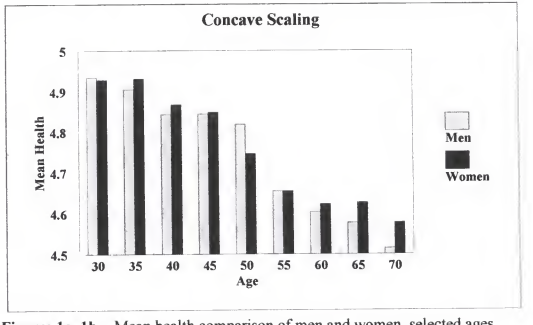
The mean is often used as an indicator of the "aggregate" level of health in a population. Relative inequality measures usually divide observations by the mean to

obtain a "normalized" distribution that serves as the measure's primary data. Normalizing by the mean facilitates comparison of distributions with different levels of societal health, aligning the distributions so that deviations from the mean are comparable. In the present context, though, the mean is highly sensitive to re-scalings, and the ranking of two distributions by the mean may not be independent of scale. Consequently, normalizing self-reported health distributions by the mean can be problematic, which in turn casts doubt on the use of standard relative inequality measures.

To illustrate the sensitivity of the mean-ordering to scale, consider distribution x=(2,2,2,2,2) and distribution y=(3,2,1,1,3). Using the "linear" scale c=[1,2,3,4,5], the mean of x is 3.0 and the mean of y is 2.9, so that $\mu(x,c) = 3.0 > 2.9 = \mu(y,c)$. However, if we re-scale the data using c'=[1,2,3,4,10], changing only the value assigned to excellent health, then the mean of x is less than the mean of y, where $\mu(x,c') = 4.0 < 4.4 = \mu(y,c')$. Clearly, the relative ordering of x and y by the mean may depend on the scale. Can such re-orderings occur in a practical setting? Consider a comparison of mean health by age for men and women using U.S. survey data from 1994.[10] The upper panel of Figure 3 reports the mean health values for men and women for nine ages using the linear scale c. The picture painted by this diagram suggests that males enjoy a higher mean health level at each of the chosen ages. In the lower panel of Figure 3, a second, highly concave scale is used to emphasize differences between lesser categories of health, resulting in a

[10]Data are taken from the nationally-representative *NHIS*. Sampling weights added by the survey have been applied. See Section 4 for a more extensive discussion of the *NHIS*.





Figures 1a, 1b. Mean health comparison of men and women, selected ages, using two different scalings (data source: 1994 *NHIS*).

reversal of the rankings of mean health across genders for seven of the nine age levels.[11] The answer to the question of which gender is in better health can depend on the numeric scale.

This does not, however, imply that all such comparisons are subject to reversals. In certain cases, a higher mean at one scaling may entail a higher mean for all scalings. Suppose that the fraction of the overall population in the worst category of health is lower (or no higher) for distribution x than distribution y. Moreover, suppose that this is true for the fraction in the lowest two categories, the lowest three categories, and so forth. Then in an obvious sense, aggregate health is greater in x than y since, for each category, x has a higher percentage of the population than y in that category or above, and a lower percentage in that category and below. It is not difficult to show that x has a higher mean than y for every possible scale c. This unambiguous ranking of distributions is well-known in other contexts as first-order dominance, and can be defined as follows:[12]

x *first order dominates* y, written x FD y, if

$$\frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{n} x_i} \leq \frac{\sum_{i=1}^{k} y_i}{\sum_{i=1}^{n} y_i} \tag{24A}$$

for every category k, or, more simply,

$$X_k \leq Y_k \tag{24B}$$

---

[11]The figure employs the scale c=[1, 4.6, 4.96, 4.996, 5], which obviously is concave in that the gap $(c_{i+1} - c_i)$ is strictly decreasing in i.

[12]This partial-sum definition of first order dominance is reminiscent of the definition of Lorenz dominance in the context of income distributions. Note that in the present context the entries in the distributions correspond to numbers of persons rather than numbers of dollars.

for all k=1,...,n, where $X_k = \sum_{i=1}^{k} x_i \Big/ \sum_{i=1}^{n} x_i$ is the cumulative share of the population in the bottom k categories of x, and Y_k is the analogous quantity for y. It is clear from the definition that this ranking of distributions is independent of scale, since it does not mention scale at all. Note that if one were to select a numeric scale, then this definition would be equivalent to the standard definition of first order dominance as a rightward (or downward) shift in the cumulative distribution function, where X_k and Y_k are the respective cumulative probabilities associated with the value c_k. The inequalities $X_k \leq Y_k$ ensure that x has greater weight on higher values, and hence a higher mean for any scale. We have the following result, whose formal verification is given in the Appendix.

Theorem III.1 One distribution, x, first order dominates another, y, if and only if the mean of x is as large as the mean of y for every scale c.

Theorem III.1 is a reinterpretation of a standard result from stochastic dominance analysis which relates first order dominance to higher expected utility for arbitrary increasing utility functions. Each re-scaling corresponds to a different increasing function, so that the mean for a specific scaling corresponds directly to expected utility. As an example of the theorem, consider distributions x=(1,4,3,6,3) and y=(2,4,4,4,3), which both contain data for a population of 17 respondents. The share of the population in the lowest category is clearly greater for y than x (Y_1=2/17 > X_1=1/17); the sum of the bottom two categories is greater for y than x (Y_2=6/17 > X_2=5/17); and so on through the fifth category, where the sums are equal. Consequently, x FD y, and so by Theorem III.1

we know that x has higher (or no lower) mean health than y for all scales.[13]

The above discussion has emphasized the potential importance of reversals in rankings of mean health due to changes in the scale. It should be noted, though, that even in the presence of first order dominance, a re-scaling can alter the relative sizes of the means and thereby affect the inequality values due to normalization by the mean. Moreover, as we shall now see, there are conceptual problems of a different sort when the measure uses the mean as a standard for evaluating deviations.

2.2 The non-robustness of the mean as a reference point

When the variable is cardinal, arbitrary transformations are not admissible and the mean is a stable reference point of the distribution. But when the variable is purely ordinal, so that any re-scaling is possible, the mean's location relative to the distribution (or any subset of observations) can vary widely with changes in the scale. For example, the mean of the distribution $x=(1,3,3,6,4)$ given the scale $c=[1,2,3,4,5]$ is $\mu(x;c)=3.53$, which is located between the seventh lowest observation and the eighth. Yet when the scale $c'=[1,2,3,4,8]$ is used, the mean becomes $\mu(x;c')=4.24$, which falls between the 13th and 14th observation out of 17. Recalling the categorical nature of the original data, it is difficult to accept the notion of a central reference value that jumps across categories and observations with such ease.[14]

---

[13]Note that if there are only two health categories there is always comparability using first order dominance. Consequently, reversals of the type discussed above require there to be at least three categories, which is typically the case with SRHS data.

[14]At the very least, it makes basic facts about a distribution difficult to report. For instance, all but one of the observations in $x'=(0,1,25,0,1)$ are above the mean when the

Now even if the sensitivity of the mean to re-scaling is not considered to be inherently objectionable, it may still pose serious problems for measuring inequality using standard tools, given their dependence on the mean as a reference point from which deviations are gauged. For example, applying the variance σ^2 to distributions x=(1,3,3,6,4) and y=(2,4,4,4,3) yields $\sigma^2(x;c) < \sigma^2(y;c)$ under the linear scale c. With the alternative scale c′=[1,2,3,4,8] the ranking becomes $\sigma^2(x;c') > \sigma^2(y;c')$. This reversal is at least partly due to the repositioning of the means of the distributions when the scale changes. For if the variance of each is evaluated at the *old* scale but using the *new* mean values, this still causes a reversal of the original ranking.[15] In other words, the non-robustness of the mean as a reference point can itself be responsible for the non-robustness in inequality comparisons as the scale is varied.

2.3 Alternative reference points

If the mean is unsuitable as a reference point in this context, then what alternative standard should be used? The above discussion has shown that the basis of the problem lies in the mean's undue sensitivity to the chosen scale. We should restrict consideration to concepts that are in some sense robust to changes in the scale. Moreover, given the ultimate use of the reference value as the standard against which deviations are to be

---

linear scale c is used; yet all but one are below it when the scale changes to c′. It is hard to imagine how one would respond to a policy-maker's question "what percentage of the population has above-average health?"

[15]In this example $\sigma^2(x;c)=1.43 < \sigma^2(y;c)=1.63$ and $\sigma^2(x;c')=5.12 < \sigma^2(y;c')=4.93$. The variances of x and y at the old scale c about the new means $\mu(x;c')$ and $\mu(y;c')$ are respectively 1.92 and 1.91.

evaluated, it should be "central" in some "space."

One possibility would be to select the health status associated with the middle category ("3" in the linear scaling) as the fixed standard of comparison. To be sure, a change in scale would have no effect on the relative position of the reference point in the distribution. And in an obvious sense, it is at the center of the set of possible health statuses. However, it is equally clear that this approach goes too far in the direction of insensitivity, since it pays absolutely no attention to the specific distribution under consideration. For example, if most of the observations in a given distribution were initially at the lowest health status, while the rest were at the second lowest, a change in the distribution that lowered all observations to the lowest status would *increase* inequality when measured as a function of the deviations from the reference point. And, yet, since all observations are now at the same level, the new distribution is clearly *more* equal. Using a fixed health status as a reference level for inequality analysis can hardly be justified.

On the other hand, "people space" has a natural center at the 50th percentile, which would suggest the use of the median health status m(x;c) as a reference level. It is clearly responsive to significant changes in the distribution – the median is always located at the position where half the population has a health status below (or equal to) it and half is above (or equal to) it. In particular, unlike a standard fixed in "category space," the median will always be the commonly held health status when the distribution is completely equal. At the same time, the median is not too responsive to changes in scale: if i is the median category in x under c, so that c_i=m(x;c), then i is the median category under any other scale c′, i.e., c'_i=m(x;c′). Unlike the mean, the relative position

of the median does not change as the scale changes. It is for these reasons that we adopt the median as our standard for evaluating inequality.

3. A median-based approach

Our approach to measuring health inequality is perhaps best introduced with the help of a simple example. Let x=(4,5,7,5,4) and y=(3,4,11,4,3) be two distributions under consideration, and take c=[1,2,3,4,5] to be the scale. Notice that 3=m(x;c)=m(y;c) is the median of both distributions. Furthermore, two persons with median health in y have lower health statuses (1 and 2) in x, while two other persons in the median category in y have higher statuses (4 and 5) in x. The remaining persons have unchanged status. Consequently, distribution x is in an intuitive sense more "spread out" than y. This relationship can be visualized with the help of a graph. Figure 4 depicts the cumulative distribution functions (CDFs) of x and y given c. Notice that the CDF for x takes on the values (.16, .36, .64, .84, 1.00) at the scale levels 1,2,3,4,5, respectively, while the cumulative values for y are (.12, .28, .72, .88, 1.00). The dashed line halfway up the vertical axis cuts both CDFs at the common median level of 3. For each status below the median, the CDF of x is above the CDF of y, while for categories above the median the opposite is true. Distribution x has greater inequality than y in that more of its mass is concentrated in the extremes, unambiguously further away from the median.

This intuitive notion of inequality as "spread away from the median" can be more rigorously defined as follows: we say that x has *greater spread* than y, written x S y, if (i) x and y have the same median category m, (ii) for all categories $k < m$, we have $X_k \geq Y_k$, and (iii) for all categories $k \geq m$ we have $X_k \leq Y_k$, where, as before, X_k and Y_k are the

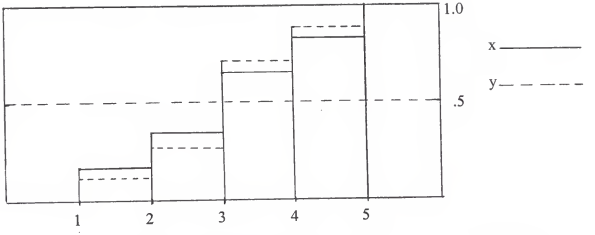


Figure 4. Cumulative distribution functions for x = (4,5,7,5,4) and y = (3,4,11,4,3).

respective cumulative population shares of the bottom k categories. In other words, y first order dominates x below the median while x first order dominates y for the median category and above -- a relationship that might be termed a "median preserving spread." As can be verified from Figure 4, this is precisely the relationship between x and y in our example. Of course, not all pairs of distributions are related to one another by a median preserving spread. Consequently, the ranking S is a *partial ordering* of distributions that is *reflexive* (x S x for any x) and *transitive* (x S y and y S z implies that x S z, for any x, y, z), but not *complete* (since for some x and y we have neither x S y nor y S x).[16] In this way it is similar to the well-known Lorenz ranking of income distributions.

The partial ordering S can be represented by a graphical tool we call the "S-curve," which is derived from the CDF as follows.[17] First, the portion of the CDF to the right of the median health status is flipped over to the left, as shown in Figure 5. The resulting curve is then rotated 90 degrees to obtain the S-curve depicted in Figure 6. Note that the base of the S-curve represents the range of the population having the median level of health. Just outside of this range are the two groups that are one category away from the median; and beyond this are the groups that are two categories from the median. Now suppose that one distribution has an S-curve that is "closer in" or "higher" than another. As we move outward from the median person along the S-curves, the one that is higher departs from the median earlier and hits every subsequent category sooner, indicating that a greater share of the population is further away from the median.

---

[16]On the use of partial orderings in related contexts, see Foster and Sen (1997).

[17]An analogous approach for income distributions was introduced by Foster and Wolfson (1993).

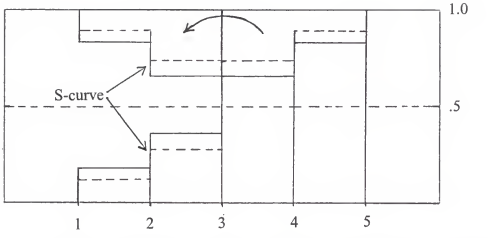


Figure 5. Deriving the S-curves from the CDF. The portions of x (solid) and y (dashed) that lie above the median are 'flipped' across the median to form the upper halves of the S-curves for x and y.

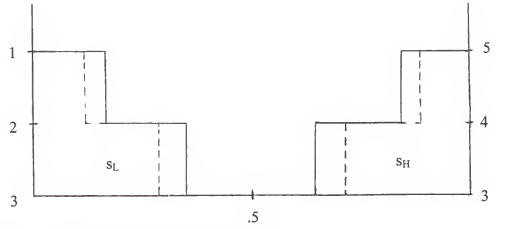


Figure 6. S-curves for x (solid) and y (dashed).

Thus a higher S-curve indicates greater inequality according to the ranking S. Conversely, if the S-curves cross, then it is immediate that the two distributions cannot be compared using S. We have the following result whose formal proof can be found in the Appendix.

Theorem III.2 Let x and y be two distributions having the same median category. Then x S y if and only if the S-curve of x is no lower than the S-curve of y.

The S-curve represents S in much the same way as the Lorenz curve represents the Lorenz ranking of income distributions.

Several useful indexes of "spread" can be derived from the S-curve analogous to the way that the Gini index is obtained from the Lorenz curve. Let $s(x;c)$ be twice the area below the S-curve of x, while $s_L(x;c)$ is twice the area to the left of 0.50 and $s_H(x;c)$ is twice the area to the right, as depicted in Figure 6. Since c is the linear scale, the height of the S-curve is the number of health categories that the associated observation departs from the median. Consequently, $s_L(x;c)$ is the average number of categories that the lower half of the distribution falls below the median, while $s_H(x;c)$ is the average number that the upper half is above the median. They are measures of the "lower spread" and "upper spread" respectively. Clearly, $s(x;c) = s_H(x;c) + s_L(x;c)$ is the average number of categories that the higher group is above the lower group, a natural if crude indicator of the overall spread of the distribution.[18] By construction, each is consistent

[18]Thus S is a median-based analogue of the Robin Hood index in income inequality analysis, which indicates how much would have to be taken from the "rich"

with the S ranking in that x S y implies $s_L(x;c) \geq s_L(y;c)$, $s_H(x;c) \geq s_H(y;c)$ and $s(x;c) \geq s(y;c)$. The converse, of course, does not hold, since the area below the curves could be larger even when curves cross.

Up to now, we have focused on the cumulative distribution functions, the S-curves, and indexes associated with the linear scale c=[1,2,3,4,5]. Our original goal was to find a method of evaluating inequality that is independent of scale. So let us select some other scale c′ and examine what happens when it is used in place of c. Clearly, the CDFs are stretched or compressed horizontally according to the new values assigned to categories, but the magnitudes of the vertical jumps is unchanged. As mentioned above, the median category is unaffected, as is the size of the population in this category, and consequently the base of the S-curve is the same as before. The rest of the S-curve is shifted in height reflecting the new distances between category values implicit in the new scale. Its height is no longer the number of health categories between a given observation and the median, but rather the number of health status units between the observation's health status and the median health status.

While the height of an S-curve is altered by a change of scale, the relative positions of S-curves are not. If the S-curve of x is higher than the S-curve of y for a given scale, this relationship remains true at all scales. Consequently the inequality ranking S corresponds exactly to the ranking implied by S-curves for any given scale. This robustness to scale ensures that the ranking S is based purely on the original data

---

(above the mean) and given to the "poor" (those below the mean), to obtain a completely equal distribution. See Atkinson and Micklewright (1992) for a description of the Robin Hood index, and Kennedy et al (1996), who use the index to analyze income inequality and health across the United States.

and not on the particular scale selected – a characteristic that is evident from the definition of S, which makes no reference to scale.

In contrast to the scale independence of the ranking S, the index s(x;c) depends crucially on c, the particular scale employed. For any c, let $\mu_L(x;c)$ and $\mu_H(x;c)$ represent the mean health status for the groups respectively below and above the median. Then $s(x;c) = \mu_H(x;c) - \mu_L(x;c)$, and hence s now measures average distance in terms of health status units rather than number of categories. The magnitude of the index surely depends on the particular scale employed. More importantly, the way that it ranks distributions can depend on the scale, since s is a difference of scale-dependent means.

Following the logic of Theorem III.1, we might ask: when can a pair of distributions be consistently ranked by the index s over *all* scales? Clearly, if x S y, then the areas under the S-curves must follow this ranking for any given scale, which is sufficient to ensure a consistent ranking by the spread index. As the following result shows, this is necessary as well.

Theorem III.3 Let x and y be two distributions having the same median category. Then x S y if and only if $s(x;c) \geq s(y;c)$ for every scale c.

Thus, the partial ordering S is equivalent to an unambiguous ranking by the spread index s at all scales.

In general, the decision of the spread measure s is influenced by the scale, and hence it may not be appropriate as a final arbiter of inequality when a single scale is used. Nevertheless, it does have a useful role as a partial signal of when S holds. If

s(x;c) is greater than or equal to s(y;c) for some c, then x S y is possible; but if s(x;c) is less than s(y;c) for some c, then x S y cannot hold.

The partial spread indexes s_L and s_H provide even greater information about the relationship between two distributions, since they separate out changes below the median from those above. Note that $s_L(x;c) = m(x;c) - \mu_L(x;c)$ and $s_H(x;c) = \mu_H(x;c) - m(x;c)$ must both follow the ranking S, when it applies. It turns out that when s_L and s_H consistently rank distributions for all scales, and are in agreement as to which distribution has greater spread, the distributions are likewise ranked by S.

Theorem III.4 Let x and y be two distributions having the same median category. Then x S y if and only if $s_L(x;c) \geq s_L(y;c)$ and $s_H(x;c) \geq s_H(y;c)$ for all scales c.

This result emphasizes the potential usefulness of s_L and s_H as signals for determining when S applies. To find S comparisons in practice, it is useful to compute s_L and s_H for the linear scaling and then to check whether s_L and s_H order the distributions in the same direction. If they do, a ranking by S is possible, and the corresponding S-curves can be further evaluated for the presence of the dominance relation S. On the other hand, if s_L and s_H disagree, a ranking by S is impossible.

Note that a disagreement between s_L and s_H suggests that one distribution has less mass than the other below the median and more above it, which is indicative of the first order dominance relation. Indeed, there is a useful relationship between s_L, s_H and FD as provided by the following result:

Theorem III.5 Let x and y be two distributions that have the same median category. Then x FD y if and only if $s_L(x;c) \leq s_L(y;c)$ and $s_H(x;c) \geq s_H(y;c)$ for all scales c.

Intuitively, when s_L unambiguously falls and s_H unambiguously rises, this ensures that the left side of the S-curve falls while the right side rises. By the relationship between the S-curve and the CDF, this is equivalent to first order stochastic dominance.

This section has presented a basic approach to evaluating the distribution of self-reported health that respects the qualitative nature of the data. Our key rankings, S and FD, render judgements only in cases where scale does not matter; otherwise they are entirely silent. This raises a natural issue of how "completely" they rank distributions in practice, which, of course, is an empirical question. The next two sections offer preliminary evidence of the applicability of our approach, and in the process offers insight on how the distribution of health varies over the life-cycle, and across states and regions of the United States.

4. Health over the life cycle

We examine the distribution of SRHS over the life cycle using data from the 1990-1994 *National Health Interview Survey (NHIS)*. The *NHIS* is a nationally representative sample of the non-institutionalized population. Approximately 44,000 households are surveyed annually, including over 100,000 individuals of all ages. As a rule, adults respond on behalf of children, while a substantial percentage (30% in 1994) of adult responses are also by proxy. Due to the complex, multi-stage sampling design

and over-sampling of certain populations, sampling weights have been appended to the survey data by the National Center for Health Statistics to render the sample representative of the U.S. non-institutionalized population. We apply these sampling weights in all of our analysis.

Using a linear scale for SRHS data from the *NHIS* and *PSID*, Deaton and Paxson (1997, 1998) find that health disperses with age.[19] The middle panel of Figure 7 reproduces their basic results for female respondents using NHIS data from 1990 through 1994. Variance appears to rise monotonically until about age 55, after which the pattern is unclear.[20] The upper and lower panels of Figure 7 illustrate the effect of changing scales. The upper panel uses a convex SRHS scale c=[1, 1.1, 1.4, 2.3, 5], where distances between upper categories are greater than distances between lower categories. Variance appears to be nearly constant until about age 40, after which variance falls. In the lower panel, the concave scale c=[1, 3.7, 4.6, 4.9, 5] is used to emphasize differences between lesser health categories, and the variance of SRHS rises nearly monotonically throughout the life-cycle. Age-to-age variance orderings, e.g., age 40 compared to age 70, are clearly dependent upon the numeric scaling of the SRHS data. Reasonable

---

[19]Deaton and Paxson's scale increases in ill-health, whereas the scale used here increases in health. Nevertheless, inequality orderings using variance or the S index introduced above are equivalent for an equidistant scale whether it increases in ill-health or health.

[20]At higher ages, there is also the question of how sample selection due to death affects trends in the dispersion of health. While it is clear that mean health would be lower if the deceased within a cohort were somehow included in the sample, it is not clear how their inclusion would affect life-cycle patterns in dispersion. The effect of excluding institutionalized individuals from the sample on dispersion is similarly unclear, though such selection likely raises the mean.

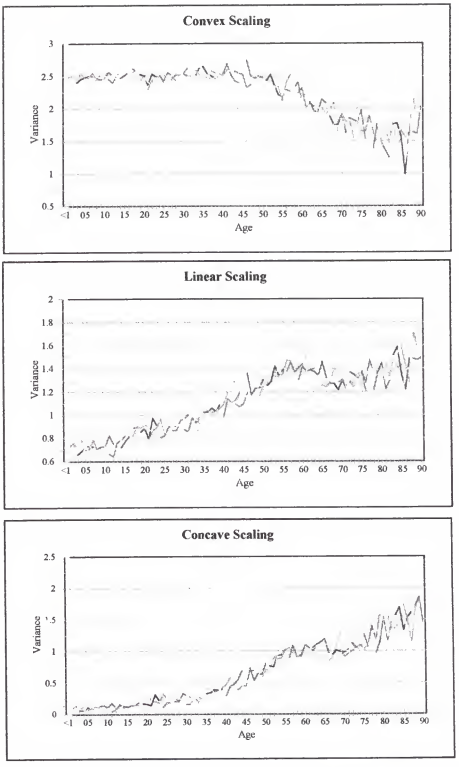


Figure 7. Variance in SRHS for Females. Data grouped by cohort, between 1990 and 1994, using three different scales (data source: 1990-1994 *NHIS*).

restrictions on the SRHS scale might narrow the results somewhat. For example, limiting SRHS to a linear or concave scaling appears to eliminate the possibility that variance *falls* over a broad range of the life-cycle. However, without such restrictions, one cannot conclude that dispersion in SRHS increases over the life-cycle.

Instead of using the variance with its obvious dependence on scale, we apply the scale independent methodology proposed above to evaluate the distribution of health for various age groups. As a first indication of potential comparisons, we calculate s_L, s_H, and the median for the linear scaling. If the medians are different for two distributions, or if s_L and s_H disagree on their ranking, then the distributions cannot be related by S. In Table 3, results from cross-sectional *NHIS* data from 1994 are reported.[21] As Table 3 shows, when the median is constant, s_H tends to fall over the life-cycle while s_L tends to rise. Hence, different ages cannot be related by S, which implies that dispersion in SRHS does not unambiguously increase (or decrease) over the life-cycle. According to Theorem III.5 above, when s_H and s_L move in opposite directions, it can be a signal of a first order relationship between the compared distributions. The clear pattern observed in Table 3 signals the possibility that ages may be related in a first order fashion: younger age groups first order dominate older age groups. Moreover, the mean health level falls with age (apart from the very old and very young) which by Theorem III.1 further indicates a first order relationship. The cumulative frequencies depicted in Figure 8 confirm that, with few exceptions, the younger ages first order dominate older ages in

[21]Using *NHIS* data from 1983-1994, Deaton and Paxson (1997) regress specific percentiles of SRHS (scaled 1,2,3,4,5) on age and cohort and find that cohort effects show little statistical significance.

Table 3. SRHS distributional indicators for females and males using the linear scale (data source: 1994 *NHIS*).

	Females					Males				
Age	**s**	**s_L**	**s_H**	**Mean**	**Median**	**s**	**s_L**	**s_H**	**Mean**	**Median**
10	1.45	1.45	0.00	4.28	Excellent	1.41	1.41	0.00	4.29	Excellent
20	1.49	0.76	0.73	3.99	Very Good	1.46	0.49	0.97	4.24	Very Good
30	1.52	0.80	0.71	3.95	Very Good	1.47	0.60	0.87	4.13	Very Good
40	1.61	0.94	0.67	3.87	Very Good	1.62	0.85	0.77	3.96	Very Good
50	1.84	1.25	0.59	3.67	Very Good	1.68	0.97	0.71	3.87	Very Good
60	2.01	0.65	1.36	3.36	Good	2.03	0.56	1.47	3.46	Good
70	1.78	0.68	1.10	3.21	Good	2.03	0.78	1.24	3.23	Good
80	1.74	0.79	0.94	3.07	Good	1.87	0.98	0.89	2.96	Good
90	1.84	0.81	1.03	3.11	Good	1.30	0.42	0.88	3.23	Good

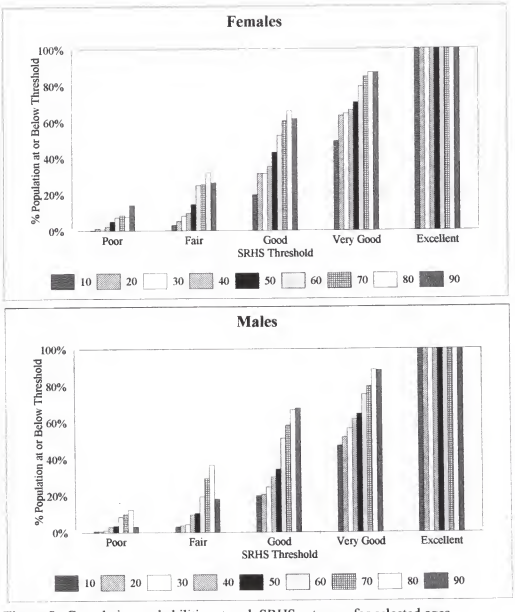


Figure 8. Cumulative probabilities at each SRHS category for selected ages (data source: 1994 *NHIS*).

health. The likelihood of being in a given health category, or below, is generally smaller at earlier ages. While the trend in the dispersion of SRHS over the life-cycle is ambiguous, there is an unambiguous decline in SRHS over nearly the entire life-cycle.

It is conceivable that part of the first-order relationship between age and health over the life-cycle could be explained by the correlation between health and income, given the covariation between age and income. As an initial test of whether the first order relationship between SRHS and age is still present after controlling for income, we compare the CDFs of neighboring 10-year age groups within seven income bands.[22] For males, we find that the FD ordering by age is substantially robust to this sort of "control" across the middle age groups. Exceptions to the FD ordering by age are concentrated in the youngest and oldest age groups, e.g., males less than the age of 10 did not generally first order dominate males between the ages of 10 and 20, and males between the ages of 10 and 20 do not generally first order dominate males in their 20s. For these younger age groups, exceptions to the age ordering of health are scattered across income categories. Among older males, exceptions to the age ordering of health are generally concentrated in the lower income categories. For females, the FD ordering by age is stronger: the health-age relationship also holds in the lowest age groups for all income bands. As with males, the health-age relationship weakens for older females. Exceptions to the age ordering of health are even more concentrated in the lower income categories for women than for men. Hence, for the *poor* elderly, the answer to the question of whether mean

---

[22]Data are a cross-section from the 1994 *NHIS*. The seven income bands: under $9,999, $10,000 to $14,999, $15,000-$24,999, $25,000-$34,999, $35,000-$49,999, and $50,000 and above. The 10-year age groups were: 0-9 years, 10-19 years, and the 20s, 30s, 40s, 50s, 60s, 70s, and 80s.

health decreases monotonically with age will depend upon the scaling of the SRHS data.[23]

5. Health across states

In this section we use our measurement techniques to obtain an initial glimpse of the distribution of health across states and regions of the United States. Much has been said about statewide (aggregate) health levels and how they vary.[24] However, relatively little is known about the distribution of health within states. This is perhaps surprising, since an evaluation of the effectiveness of publicly-funded health programs (such as Medicaid) would seem to require analysis of this type.[25] In any event, it seems reasonable to assume that policy makers would be better able to target health interventions if they had a clearer picture of the distribution of health needs. Hence, a principal motivation for evaluating the distribution of health across states is to provide information that might be useful to those who must decide upon the future course of publicly-funded health programs.

---

[23]Another way to express this finding is that health poverty, measured against various SRHS categorical thresholds, does not necessarily increase continuously with age among the poor elderly. Once again, it is unclear how sample selection due to death and institutionalization affects these findings.

[24]See for example Kennedy, Kawachi and Prothrow-Stith (1996) and Kaplan et al (1996), who relate income inequality to state health levels, and Kennedy et al (1999), who examine differences in the health status of men and women.

[25]Medicaid, for example, facilitates large intra- and inter-state redistributions of resources, targeting aid to many of those with the greatest health needs, including the aged, the poor, the disabled, and those with large medical bills.

We use SRHS data from the 1994 *NHIS State Data Files*[26] to compare the distribution of health across states, checking for both first order and S relationships. Each state's distribution has been standardized to mimic the age and sex composition of the U.S. as a whole.[27] Following the diagnostic strategy employed in our analysis of health over the life-cycle, we first calculate means, medians, the s index, and its components, s_H and s_L, using the linear scaling for SRHS distributions from the 25 largest states.[28] Results from the adjusted sample are reported in Table 4. Notice first that all selected states have "very good" as their median health category. When the linear scaling is used, the index s can be interpreted as the average number of categories separating those in the upper and lower halves of state health distributions. The average categorical spread between the upper and lower halves of the distribution ranges from about 1.5 in Minnesota and Wisconsin to nearly 2.0 in Alabama. In general, southern

---

[26]The *NHIS State Data Files* append state identifiers to standard *NHIS* data. To maintain the confidentiality of respondents, variables from the *NHIS* that are included in the *NHIS State Data Files* are aggregated, e.g., continuous variables are categorized. Several small states are grouped together due to small sample sizes, and two states, Nebraska and North Dakota, are not sampled by the *NHIS*.

[27]A direct method of standardization was employed. For any given state k=1...25, the adjusted number of people in each SRHS category i is given by the equation $SRHS_{ki}^{A} = \Sigma_{d=1}^{D}\left[SRHS_{kdi} / n_k\right] * \left[\pi_{kd} / \pi_d\right] * n_k$, or $SRHS_{ki}^{A} = \Sigma_{d=1}^{D} SRHS_{kdi} * \left[\pi_{kd} / \pi_d\right]$, where $SRHS_{kdi}$ is the unadjusted number of people in state k and age/sex category d in SRHS category i, π_{kd} is the proportion of state k's population in age/sex category d, and π_d is the same proportion for the nation as a whole. Since the *NHIS State Data Files* report age in six categories, D = 6 * 2 = 12.

[28]The states selected are the 25 largest, according to the population estimates included in the 1994 *NHIS-SDF*, and comprise approximately 85% of the total US population. We included only the largest states in our analysis in order to minimize the effects of sampling error due to the potential non-representativeness of the state sub-samples in the *NHIS*.

Table 4. SRHS distributional indicators for the 25 largest states using the linear scale (data source: 1994 *NHIS State Data Files*).

		s	s_L	s_{II}	Mean	Median
Northeast						
	Massachusetts	1.65	0.75	0.90	4.08	Very Good
	New York	1.58	0.83	0.74	3.96	Very Good
	New Jersey	1.66	0.95	0.72	3.89	Very Good
	Pennsylvania	1.68	0.91	0.77	3.93	Very Good
Midwest						
	Ohio	1.68	0.90	0.78	3.94	Very Good
	Indiana	1.66	0.97	0.69	3.86	Very Good
	Illinois	1.62	0.86	0.77	3.95	Very Good
	Michigan	1.74	1.02	0.72	3.85	Very Good
	Wisconsin	1.57	0.84	0.73	3.95	Very Good
	Minnesota	1.54	0.69	0.84	4.08	Very Good
	Missouri	1.69	0.99	0.70	3.85	Very Good
West						
	California	1.74	1.03	0.71	3.84	Very Good
	Washington	1.58	0.77	0.81	4.02	Very Good
	Colorado	1.71	0.81	0.90	4.04	Very Good
	Arizona	1.84	1.10	0.73	3.82	Very Good
South						
	Maryland	1.58	0.76	0.82	4.03	Very Good
	Virginia	1.66	0.94	0.72	3.89	Very Good
	North Carolina	1.87	1.21	0.66	3.73	Very Good
	Georgia	1.81	1.02	0.79	3.88	Very Good
	Florida	1.77	1.01	0.76	3.87	Very Good
	Alabama	1.96	1.35	0.61	3.63	Very Good
	Tennessee	1.80	1.12	0.68	3.78	Very Good
	Louisiana	1.87	1.10	0.77	3.83	Very Good
	Texas	1.80	1.03	0.77	3.87	Very Good
	Oklahoma	1.72	0.97	0.75	3.89	Very Good

states have a wider categorical spread in their distributions of health. Southern states also tend to have lower levels of mean health using the linear scale. Ad hoc state-to-state comparisons of s_L and s_H suggest that it is more typical for the two to disagree. This signals the possibility that, where present, dominance relationships are more likely to be characterized by the first order relation than the S relation.

The results of a comprehensive search for first order and S relations among the 25 largest states are presented in Tables 5 and 6. If a cell in Table 5 contains a "1" then the column state has more overall health than the row state. If a cell in Table 6 contains a "1" then the column state has a more equal distribution of health than the row state. The states are grouped by census region to help identify any regional patterns in the distribution of health. In this selection of states, we find that comparisons can be made in 254 of the possible 300 state pairs, with 199 of these being first order relationships and the remaining 55 being S rankings. Based on the results presented in Table 5 it is clear that southern states tend to be first order dominated by northern, mid-western, and western states.[29] Thus, applying Theorem III.1, southern states have the least health on a per-capita basis, independent of any numeric scaling of the SRHS data. These results are generally consistent with regional orderings by mortality rates, and imply that regional health differences in the U.S. are not limited to the lower tail of the distribution, but persist across the population. The inequality comparisons presented in Table 6 do not suggest a clear regional pattern. Nevertheless, where S relationships can be established,

[29]In fact, a comparison of health across census regions using data from all states included in the 1994 *NHIS* found a complete first order ordering as follows: the Northeast had the most overall health, followed by the Midwest, the West, and the South.

Table 5. First order dominance in SRHS among the 25 largest states. A '1' in a cell indicates that the column state has more overall health than the row state, i.e., that the column state first order dominates the row state (data source: 1994 *NHIS State Data Files*).

	Northeast				Midwest							West				South										
	MA	NY	NJ	PA	OH	IN	IL	MI	WI	MN	MO	CA	WA	CO	AZ	MD	VA	NC	GA	FL	AL	TN	LA	TX	OK	Totals
Northeast																										
MA		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NY	1		0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	4
NJ	1	1		0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	7
PA	1	0	0		1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	5
Midwest																										
OH	1	0	0	0		0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	4
IN	1	1	1	1	1		1	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	11
IL	1	0	0	0	0	0		0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	4
MI	1	1	0	1	1	0	1		1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	11
WI	1	0	0	0	0	0	0	0		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
MN	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MO	1	1	1	1	1	0	1	0	1	1		0	1	1	0	1	1	0	0	0	0	0	0	0	0	12
West																										
CA	1	1	1	1	1	0	1	0	1	1	0		1	1	0	1	1	0	0	0	0	0	0	0	0	12
WA	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	1
CO	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0
AZ	1	1	0	1	1	0	1	0	0	1	0	0	1	1		1	0	0	0	1	1	0	0	0	1	1	13
South																										
MD	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	1
VA	1	1	0	1	1	0	1	0	1	1	0	0	1	1	0	1		0	0	0	0	0	0	0	0	10
NC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	0	1	1	1	1	23
GA	1	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0	0		0	0	0	0	0	0	5
FL	1	0	0	1	1	0	1	0	0	1	0	0	1	1	0	1	0	0	0		0	0	0	0	0	8
AL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	24
TN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0		0	1	1	19
LA	1	0	0	1	1	0	0	0	0	1	0	0	1	1	0	1	0	0	0	1	0	0		1	0	9
TX	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0		0	6
OK	1	0	0	1	1	0	1	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0		8
Totals	20	10	6	12	15	3	11	3	9	22	3	3	19	14	2	19	5	1	4	4	0	2	2	5	5	199

Table 6. S dominance in SRHS among the 25 largest states. A '1' in a cell indicates that the column state has more equal health than the row state, i.e., that the row state S dominates the column state (data source: 1994 *NHIS State Data Files*).

	Northeast				Midwest							West				South										
	MA	NY	NJ	PA	OH	IN	IL	MI	WI	MN	MO	CA	WA	CO	AZ	MD	VA	NC	GA	FL	AL	TN	LA	TX	OK	Totals
Northeast																										
MA		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NY	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NJ	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PA	0	1	0		0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Midwest																										
OH	0	1	0	0		0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
IN	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IL	0	1	0	0	0	0		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
MI	0	0	1	0	0	0	0		0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	3
WI	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MN	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MO	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
West																										
CA	0	0	0	0	0	1	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0	0	0	2
WA	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0
CO	0	0	0	0	0	0	0	0	0	1	0	0	1		0	1	0	0	0	0	0	0	0	0	0	3
AZ	0	0	1	0	0	1	0	1	1	0	1	1	0	0		0	1	0	0	0	0	0	0	0	0	7
South																										
MD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
VA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0
NC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0
GA	0	1	0	1	1	0	1	0	1	0	0	0	0	0	0	0	1	0		0	0	0	0	0	1	7
FL	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0		0	0	0	0	0	5
AL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0
TN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0
LA	0	1	1	0	0	1	1	1	1	0	1	0	0	0	0	0	1	0	0	1	0	0		0	1	10
TX	0	1	1	1	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0		0	8
OK	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		2
Totals	0	8	5	2	1	5	4	2	9	2	5	1	1	0	0	1	6	0	0	1	0	0	0	0	2	55

it appears that northern and mid-western states tend to have a more equal distribution of health than southern states.

Interestingly, the above results have important implications for comparisons of health *poverty* as well. Note that a first order relationship between two states ensures that the dominant state has the same or a smaller health poverty rate irrespective of the choice of poverty threshold.[30] Similarly, an S ordering between two states ensures that the more equal state has the same or a smaller poverty rate for all poverty thresholds below the median. If we accept that the poverty threshold in this data should fall below "very good" (the universal median), then unambiguous poverty comparisons of this type are evident in 254 of the 300 possible comparisons. A "1" in a particular cell of either Table 5 or 6 implies that the column state has a lower (or the same) level of health poverty than the row state no matter the poverty line. Therefore, the regional patterns observed in the first order and inequality comparisons above are that much more pronounced when the focus is on poverty.

6.Conclusion

We have presented a median-based partial ordering that indicates when there is an increase in inequality in a qualitative distribution of health. We have also introduced in this context first order dominance, which yields unambiguous orderings of aggregate or mean health. In addition, we defined two partial indices of dispersion s_L and s_H, and

[30]The health poverty rate is defined as the proportion of a population whose health status is no greater than the minimum acceptable health category, or the poverty threshold. This approach to robust poverty comparisons was initiated by Foster and Shorrocks (1988) in the context of income distributions.

their aggregate s, and have shown how these indices can act as signals for the two orderings and how they can provide an intuitive measure of categorical spread when used in conjunction with the linear scaling. Our empirical applications have shown the prevalence of first-order relationships in the distribution of health across various sub-populations, including age and state of residence. Moreover, we have obtained preliminary results from a partial inequality ranking that suggest the south may have not only less aggregate health than the rest of the country, but that health may also be less equally distributed in the South.

A potential drawback of our approach is that inequality comparisons are limited to situations in which medians coincide. This can prove to be restrictive, particularly when first order effects are in evidence, causing medians to vary across groups, as illustrated by our analysis of health over the life cycle. Nevertheless, a great many of the evaluations do, in fact, involve common median comparisons, and in particular *all* of the state-wise comparisons presented above have this characteristic. Thus, this requirement need not be widely restrictive. Other limitations can arise from our reliance on rankings that are inherently incomplete. When two distributions can be compared using the unambiguous partial ordering S, the information conveyed is quite powerful. But when the high standards underlying the ranking fail to be achieved, what can be done? Clearly, one approach is to relax the standards somewhat by limiting the class of allowable scales. For example, one could initially rank the distributions using the index s with the linear scale, and then check for robustness of this ranking over all concave scales.[31] This would

[31]Wagstaff and Van Doorslaer (1994) and Deaton and Paxson (1997, 1998) discuss the potential importance of concavity in this context.

in principle yield a "more complete," but still partial, ordering of distributions, which would convey a somewhat less general notion of robustness. It would be interesting to explore whether such a criterion could be developed and, if so, how much additional bite it might have.

Taking this limited scale approach to the extreme, we return to the case of a single scale and hence a complete ranking of distributions. Unfortunately, we also regain scale dependence. We should note, though, that the combination of the s index (or s_L or s_H) and the linear scale c has an additional attraction that other combinations may not, namely, the value s(x;c) indicates how many *categories* separate observations below the median from those above the median, on average. The value s(x;c) has a concrete and easy to interpret meaning that is useful in describing a health distribution, and the same is true for $s_L(x;c)$ and $s_H(x;c)$. Moreover, given this inherent significance, one may with some justification explore the possibility of decomposing s(x;c) across population subgroups having the same median. In fact, it is easy to show that for a distribution x broken down into subgroup distributions $x^1,..., x^m$, having the same median, the overall spread s(x;c) is the weighted average of the group-wise spreads, $s(x^1;c),...,s(x^m;c)$, with weights given by the respective subgroup population shares. This decomposition may prove to be of some use in identifying the source of inequality in a population.

Finally, note that while the above discussion has focused on SRHS data, it is easy to see that the same methodology would be applicable to other categorical variables. For example, subjective measures of happiness[32] have properties quite similar to the health

---

[32] For example, Jencks and Torrey (1988) analyze responses to the following question from the U.S.'s General Social Survey: "Taken all together, how would you say

variable we employ, suggesting that our techniques can be used to assess their distribution. On the other hand, there appears to be no direct way of applying our methods when the variable is classified into different numbers of categories in the two distributions being compared.[33] It remains to be seen whether our approach can be extended to cases like this.

---

things are these days -- would you say that you are very happy, pretty happy, or not too happy?"

[33]This question is raised by Wagstaff and Van Doorslaer (1994) in a somewhat different context. Their solution is not applicable to the problem we are addressing. In particular, their method is based on the assumption that all countries have the same latent distribution of health.

CHAPTER IV

THE DISTRIBUTION OF SELF-REPORTED HEALTH IN THE UNITED STATES: A RESTRICTED SCALE APPROACH

1. Introduction

The distributional rankings generated for qualitative SRHS data in chapter III rely solely on the ordinal quality of the data and, as a result, are entirely independent of the choice of numeric scale, so long as the scale is strictly increasing in good health. To compare overall -- or mean -- levels of health in two distributions, one calculates the two cumulative distribution functions (CDFs): if the CDFs are unequal and do not cross then the mean ordering of the two distributions is independent of scale. This circumstance describes a first order dominant (FD) relationship between the two distributions. If, for example, x FD y, then x has a higher mean level of health than y for every scale. Equivalently, if different scales produce different mean orderings of two distributions, then they cannot be related by the FD relation. The median-based inequality comparisons proposed in chapter III work in a similar fashion. A scale-independent S inequality ordering exists when the S-curves for two distributions do not cross, or equivalently, when the CDFs for two distributions with a common median category cross exactly once at the median.[1] This circumstance exists when the FD ordering for the upper and lower halves of two distributions are different. When x S-dominates y then, no

[1]The latter description of S dominance presumes an interior median category. When the median is the lowest or highest category, the CDFs will not cross, and the S and FD orderings will coincide.

matter what scale is used, x will have both a larger upper mean and a smaller lower mean, and will thus be more unequal, than y. These descriptions emphasize the manner in which FD and S can be thought of as, respectively, mean and partial-mean orderings that are entirely robust to scale.

Despite the seemingly heavy burden of proof required to generate these scale-independent distributional comparisons, a large number of FD and S orderings are found among states. Using 1994 data to compare each of the 25 largest states, two at a time, the methodology proposed in chapter III identifies first order relationships in 199 of the 300 distinct pairs of states. The clear pattern that emerges from these comparisons is that overall health is less in southern states than in other parts of the country. Somewhat fewer state-to-state inequality relationships are observed: 55 of the remaining 101 state pairs could be compared using the scale-independent S relation. These comparisons suggest a similar pattern, i.e., that health is less equally dispersed in the south than in other parts of the country. As described above, these results are entirely robust to changes in the scaling of the SRHS data, but since a complete distributional ordering of states is not obtained, the question remains whether or not the results are representative of the true patterns in the distribution of health across the United States.

The most convincing way to address the question of representativeness would be to generate distributional comparisons where the FD and S relations do not apply. It is useful, then, to look more closely at the circumstances that lead to failures of the FD and S relations. If 199 of 300 state pairs can be compared using the FD relation, as above, then the mean ordering in the remaining 101 state pairs must be a function of scale: for these 101 state pairs, mean *re*-orderings can occur when certain scales are used.

Similarly, if 55 state-pair comparisons are characterized by the S relation, then it must be the case that in the 245 remaining state-pair comparisons, the inequality (partial mean) ordering of the two states depends on the choice of numeric scale for the SRHS data. Our understanding of the distributional relationship between states that cannot be related by the FD or S relations might be enhanced if we knew something about the scales that caused the mean or partial-mean re-orderings of these state-pairs. If, for example, mean health in state x were greater than mean health in state y for all but a narrow range of scales, then it would be helpful to know whether there was any justification for the use of those scales. Identifying *unreasonable* scales for which no justification exists would enable the exclusion of mean re-orderings associated with those scales. If, after excluding unreasonable scales, all remaining scales generated consistent mean orderings, then the distributions could presumably be compared.

To illustrate the kinds of exceptional orderings that might be disregarded because they are generated by unreasonable scales, consider the following comparison of SRHS distributions x and y. As seen in the upper panel of Figure 9, x first order dominates y through the third category, but y has the larger percentage of its population in the first four categories combined. Recalling that the mean of a distribution is the area above its CDF, it appears by visual inspection that x has a higher mean than y for the scaling used in the upper panel of Figure 9, $c=[c_1, c_2, c_3, c_4, c_5]$. Nevertheless, when the distance between the fourth and fifth categories is greatly exaggerated, as with the scale $c'=[c_1', c_2', c_3', c_4', c_5']$ shown in the lower panel of Figure 9, it seems that the mean ordering of x and y has been reversed. Even though there may be no justification for the use of c', its existence rules out comparisons of overall health between x and y using the first order

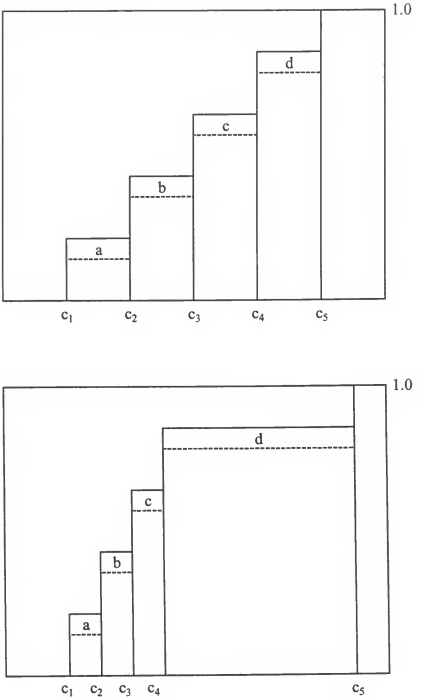


Figure 9. Mean ordering of x (dashed) and y (solid) under two different scales. In the upper panel, the mean of x is greater than the mean of y, since a+b+c>d. In the lower panel, the mean of x is less than the mean of y, since a+b+c<d.

criteria, since the first order criteria requires unanimity in mean orderings over all scales.

This discussion leads to two important questions. First, is there some way to identify which scales are reasonable and which are unreasonable? Second, if a set of reasonable scales can be identified, then is there some way to check for consistency in the ordering of means and partial-means among this restricted set of scales? In the remaining sections of this paper, these questions are addressed, and a methodology is introduced that successfully generates new distributional comparisons among states above and beyond those identified by the FD and S relations. The paper is organized as follows. Section 2 presents a rationale for restricting the set of scales applicable to SRHS data to those that are concave, i.e., scales that emphasize distances between lesser categories. Section 3 introduces a methodology for generating distributional rankings that apply to this limited set of applicable scales. The methodology is shown to support comparisons of overall health, health inequality, and health poverty. In section 4, these methods are applied to data from the 1994 *NHIS-SDF* to characterize interstate and interregional differences in the distribution of health. Section 5 summarizes the methodological and empirical results.

2. Restricting the SRHS scale

For certain qualitative variables, we may be able to infer something about the relative distances separating categories. The categorical labels themselves may impart some information regarding these relative distances. As an extreme example, consider a hypothetical variable that measures respondents' emotional state. If the available categories are "really extremely sad, happy, happier than normal, and ecstatic," it might

reasonably be assumed that the largest distance between neighboring categories may be found between the really extremely sad and happy categories, and that the least distance exists between the happy and happier than normal categories. For the SRHS variable used in the *NHIS*, the response categories are "poor, fair, good, very good, and excellent." Though many such restrictions on the potential scaling of this variable can be imagined, a particularly strong argument can be made that the scale should be concave, i.e., that distances between lesser health categories should be greater than or equal to distances between greater health categories.[2]

Grossman (1972) suggests that individual health can be represented in a single dimension as a stock, with positive and negative flows determined primarily through rates of investment and depreciation. Individuals choose their level of health through, for example, investments in time and medical care, while age is a primary determinant of the rate of depreciation in health. In this model, one's stock of health yields returns in the form of healthy days of life. Grossman interprets death as the result of the (variously constrained) choice of a deficient stock of health. Consistent with this view of overall health status, Figure 10 below depicts the various categories of self-reported health along a uni-dimensional scale at hypothetical distances above some minimum threshold representing death. The distances between categories in Figure 10 decline as one moves up the scale, an interpretation that seems appropriate given the labeling of the categories.

---

[2]Precisely, $(c_{j+1} - c_j) > (c_j - c_{j-1})$ for all interior categories j. Other restrictions on the scale of SRHS data could be explored, e.g., also excluding certain classes of concave scales. Nevertheless, the rationale offered here would not appear to support further scale restrictions. Moreover, the simple assumption of concavity facilitates a significant increase in measurement power, as described in section 3 below.

Figure 10. Various levels of health in relation to the minimum threshold where death occurs.

The distinction between very good and excellent, for example, is surely finer than the distinction between poor and fair. Some indication of respondents' interpretation of the relative location of SRHS categories may be obtained from the work of Gerdtham et al (1999), who use data from a Swedish survey that asks respondents to rate their health both categorically and numerically. Gerdtham et al show that respondents who place themselves in higher health categories also tend to assign themselves higher numeric health scores (on a 0-100 scale). Their results also indicate that the difference between the average numeric health scores for respondents in neighboring categories generally decreases as categorical health status improves.

Research examining the relationship between self-reports of general health status and subsequent mortality would appear to confirm the scale interpretation of SRHS offered above. Though exact mortality rates vary from sample to sample, available research strongly suggests that the difference in mortality rates (and, presumably, levels of health) for those with poor versus fair SRHS is greater than the difference in mortality rates (and health levels) for those with fair versus good SRHS, and that these inter-category differentials continue to decline as categorical health status improves.[3] The

[3]See for example Mossey and Shapiro (1982), Idler and Angel (1990), Idler and Kasl (1991), Grant, Piotrowski and Chappell (1995), and Appels et al (1996).

relationship between mortality and SRHS undergirds Deaton and Paxson's (1998) argument that, when calculating aggregate social welfare, individual utility levels should be represented by some concave transformation of SRHS.

If any reasonable scaling of the SRHS variable is concave in good health, then non-concave scales can be excluded from consideration in distributional comparisons, which is to say that any mean or partial mean re-orderings associated with strictly convex scales can be ignored. The question then becomes whether a method exists for comparing distributions that takes advantage of the scale limitations proposed above, i.e., that identifies distributional relationships that are robust to reasonable (i.e., concave) scalings, but that are not necessarily robust to unreasonable (i.e., convex) scalings.

3. Second order distributional comparisons

The basic result that enables us to take advantage of the concavity restriction placed upon SRHS data is borrowed from the stochastic dominance literature. A widely known result entails the use of second order dominance to establish distributional orderings. An early version of this result is presented by Hadar and Russell (1969), who show that if distribution x second order dominates distribution y, then the expected utility gained from distribution x is at least as great as that obtained from distribution y as long as preferences defined over the distributions are weakly concave.[4] A similar result is presented below, except that the concave utility functions are interpreted as concave

[4]Deaton and Paxson (1998) appeal to this result in suggesting that second order stochastic dominance may be used to establish consistent (or robust) welfare orderings between two SRHS distributions, so long as aggregate, or social, welfare is defined as an integral over some concave function of health status.

transformations (or re-scalings) of the underlying qualitative data, enabling a comparison of means, rather than a comparison of expected utilities. Note that the mean comparisons established by second order dominance apply to *any* concave transformation of the data. The robustness of mean comparisons to concave transformations of the data is especially important in the present context, as it facilitates distributional comparisons that apply to all admissible, i.e., concave, scalings of the SRHS data, but that do not necessarily apply to inadmissible scales. The exposition below describes how second order dominance may be used to establish robust mean orderings of SRHS distributions, and how this basic result can be adapted to facilitate inequality and poverty orderings as well.

3.1 Comparing overall levels of health using second order dominance

The inequality condition for second order dominance is typically obtained by integrating the pair of cumulative distributions that are to be compared. In the familiar expression below, distribution F second order dominates distribution G

$$\int_{-\infty}^{x} F(t)dt \leq \int_{-\infty}^{x} G(t)dt \text{ for all x.}^{5} \qquad (25)$$

Consider, though, SRHS distributions x and y that have been ordered from the lowest level of health to the highest, where each distribution is represented by a vector, e.g., $x=(x_1, x_2,..., x_n)$, x_j is the number of observations in category j, and n is the number of categories. For categorical data, an alternative characterization of second order

[5]In other contexts, a strict inequality for some x is required. See for example Hanoch and Levy (1969).

dominance is required. Let a scale be defined as a positive vector $c=[c_1,c_2, \ldots,c_n]$ that assigns category i the numeric value c_i, where $c_i > c_{i-1}$ for all i. Distribution x *second order dominates* y for scale c, written x SD(c) y, if

$$\sum_{j=1}^{k}(c_{j+1}-c_j)X_j \leq \sum_{j=1}^{k}(c_{j+1}-c_j)Y_j \text{ for all } k=1,\ldots,n-1, \tag{26}$$

where $X_j = \sum_{i=1}^{j} x_i / \sum_{i=1}^{n} x_i$ represents the cumulative population share for distribution x at each category j, j=1,...,n, and the Y_j's represent the corresponding population shares for distribution y. When x SD(c) y, the area under the CDF of x is less than the area under the CDF of y for any equivalent range of the two distributions that begins at the minimum level of health, c_1. This condition can also be stated in terms of the areas *above* x and y, which represent the means for the two distributions when k=n-1: when x second order dominates y the area above the CDF of x is greater than the area above the CDF of y up through category n-1, so that the mean of x is greater than or equal to the mean of y, $\mu(x;c) \geq \mu(y;c)$. The useful aspect of second order dominance when comparing distributions of SRHS is that if (26) holds for scale c, it also holds for any scale that is more concave than c.[6] This leads to a well-known result from the stochastic dominance literature that takes the following form when applied to qualitative data[7]

Theorem IV.1 One distribution, x, second order dominates another, y, for some scale c if and only if $\mu(x;c') \geq \mu(y;c')$ for every scale c′ that is a concave transform

---

[6]See the appendix for a more precise definition of concavity and for a proof of this intermediate result.

[7]See Hadar and Russell (1969) for an alternative characterization and proof.

of c.

[see Appendix for proof]

Theorem IV.1 suggests that if we find a scale c for which the areas under two distributions are consistently ordered for every range where the lower bound is the minimum level of health, then the mean will rank the two distributions consistently when any scale c′ that is at least as concave as c is used. The initial scale c that is used to establish second order dominance is referred to as the "reference" scale. If a linear, or equidistant,[8] reference scale is used to establish second order dominance, e.g., c=[1,2,3,4,5], then the resulting mean orderings will hold true for all concave scalings of the data. Hence, the SD(c) relation yields a ranking of distributions by their overall levels of health that is robust to all reasonable scalings of the data.[9] Returning to the example illustrated in Figure 9, note that the SD(c) relation enables a comparison of overall health in distributions x and y since it ignores the mean re-orderings caused by convex scales such as c′. Note that the use of the familiar c=[1,2,3,4,5] linear scale to establish second order dominance over all concave scales is innocuous: *any* positive equidistant scale may be used to establish this result.[10] Nevertheless, using the

---

[8]Categories are equidistant, i.e., distances between neighboring categories are all the same.

[9]Not all concave scalings are necessarily reasonable. Rather, it is argued, any scale that is reasonable must be concave.

[10] In particular, if $\mu(x;c) > \mu(y;c)$ for some positive equidistant scale c, then $\mu(x;c') > \mu(y;c')$ for any positive equidistant scale c′. To see this, note first that c′ can be expressed as a linear transformation of c, $c_j' = a + b*c_j$ for j=1,...,n, where b is positive constant and $a + c_j > 0$. The additive factor, a, has no effect on aggregate comparisons of health as described in (26): since each scale value is increased (or decreased) by the

"canonical" c=[1,2,3,4,5] linear scale has the advantage that the resulting population means, means for the upper and lower halves of the distribution, and the spread indices s, s_H and s_L are all presented in categorical units, so that s, for example, is equal to categorical spread.[11]

3.2 Second order inequality comparisons

Chapter III introduces the S inequality relation and advocates its use for qualitative data such as SRHS because: (1) it characterizes inequality in terms of distance from the median, which is an especially appropriate reference point for qualitative data; and (2) the orderings that S generates hold for every increasing scale c that could be applied to the data. To generate inequality orderings that are robust to concave -- though not necessarily to convex -- scalings of the data, we now turn to the partial spread indices, s_L and s_H, which chapter III defines as, respectively, twice the areas under the lower and upper portions of the S curve. When s_L and s_H agree in their ordering of two distributions for every scale, an S-ordering is generated, e.g., if distributions x and y have a common median, and if $s_L(x;c) \leq s_L(y;c)$ and $s_H(x;c) \leq s_H(y;c)$ for every c, then y S x, and x is thus found to be the more equal distribution. Note that s_L is also equal to the median minus the lower mean, while s_H is equal to the upper mean minus the median. Since the upper and lower spread indices can be expressed in terms of their respective means, and

---

same amount, the size of each inter-category gap, $c_{j+1} - c_j$, is unaffected. Also, multiplying scale c by some constant b has the same effect on both sides of the inequality in (26), and hence does not affect comparisons of aggregate health in two distributions.

[11]See section 3 of Chapter III.

a common median, it is useful to explore whether the second order results presented above could be used to generate orderings that apply for every concave scaling of the SRHS data.

First, observe that for two distributions x and y with a common median category m, x second order dominates y up to m using scale c whenever

$$\sum_{j=1}^{k}(c_{j+1}-c_j)X_j \le \sum_{j=1}^{k}(c_{j+1}-c_j)Y_j \quad \text{for all k=1,...,m-1.} \tag{27}$$

Just as equation (26) implies that x second order dominates y for c and all scales more concave than c over the entire range of distributions x and y, (27) implies that x second order dominates y for c and all scales more concave than c below the median. When (27) applies, the lower mean of x must be greater than the lower mean of y for scale c and for every scale more concave than c. Since s_L is equal to the median minus the lower mean, (27) implies that $s_L(x;c')$ must be less than $s_L(y;c')$ for all every scale c' that is more concave than c.

Theorem IV.2 Given two distributions x and y with a common median, x second order dominates y up to the median for scale c if and only if $s_L(x;c') \le s_L(y;c')$ for every scale c' more concave than c.

[see Appendix for proof]

When x SD(c) y below the median, Theorem IV.2 implies that there is less spread in the lower half of x than in the lower half of y for every scale that is at least as concave as c. If the reference scale c is linear, then this result implies that there is less categorical

spread below the median in x than in this lower portion of y, and that there is also less spread below the median in x for all *reasonable* scalings of the SRHS data.

The upper portions of the distribution can be compared in a similar fashion. Since s_H is equal to the upper mean minus the median, y second order dominates x above the median using scale c whenever

$$\sum_{j=m}^{n-1}(c_{j+1}-c_j)Y_j \le \sum_{j=m}^{n-1}(c_{j+1}-c_j)X_j \text{ for all k=m,...,n-1.} \quad (28)$$

As when comparing distributions over the entire range or below the median, when a second order relationship exists between two distributions above the median for a given scale c, the same relationship holds for all scales that are more concave than c. When equation (28) holds, the mean of the upper portion of distribution y must be greater than the mean of the upper portion of x for c and any scale more concave than c. When the upper means are so ordered, then $s_H(y;c)$ must be greater than $s_H(x;c)$ for c and all scales more concave than c, since s_H is equal to the upper mean minus the median.

Theorem IV.3 Given two distributions x and y with a common median, y second order dominates x above the median for scale c if and only if $s_H(x;c') \le s_H(y;c')$ for every scale c' more concave than c.

[see Appendix for proof]

If c is a linear scale, then Theorem IV.3 indicates that there is less categorical spread in the upper portion of x than in the upper portion of y, and that there is also less spread in the upper portion of x when alternative, yet reasonable scalings of the data are used.

If equations (27) and (28) both apply for a pair of distributions x and y using scale c, then the means of the upper and lower portions of x are closer to the median, and closer to each other, than the means of the upper and lower portions of y. This comparison of upper and lower means is captured in the definition of the second order S inequality relation.

Definition Given two distributions x and y with a common median, x *second order S dominates* y for scale c, written x S(c) y, if x second order dominates y below the median using scale c and y second order dominates x above the median using scale c.

According to this definition, Theorems IV.2 and IV.3 must be applicable in order to achieve second order S dominance. Hence, x S(c) y whenever there is less spread in x than in y in both the upper and lower portions of the distributions for every scale that is a concave transform of c. If c is the linear scale, then it follows directly that x S(c) y for all reasonable scalings of the data. It is in this sense that x may be considered to be more equal than y.

The use of second order dominance in inequality measurement facilitates comparison of distributions such as x and y, as illustrated in Figure 11. Since the CDFs for x and y cross below and above the median, they are incomparable by the first order S relation. But if x second order dominates y below the median using a linear scaling, and vice-versa above the median, then we know that dispersion in x is less than in y for all reasonable scalings of the data. Apparently, the scales that might reverse the inequality

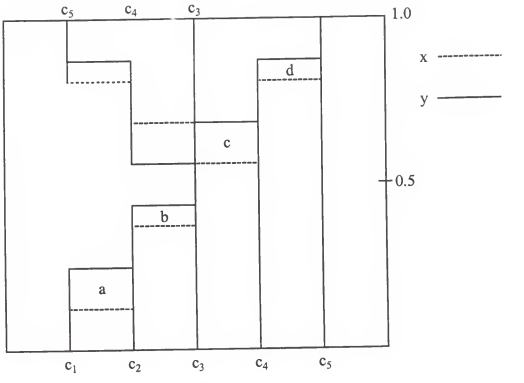


Figure 11. Second order S inequality. Since area a>b and c>d, then $s_L(x;c) < s_L(j;c)$ and $s_H(x;c) < s_H(y;c)$. Hence, y s(c') x for every scale c' that is a concave transform of c.

ordering of x and y are unreasonable, i.e., convex[12].

Note that since the partial spread indices s_L and s_H must rank x and y in the same way (and for all concave scales) if x S(c) y, the dominance aspects of the S(c) relation may be extended beyond scale. When x S(c) y, the inequality ordering is the same for any measure $I(s_L, s_H) = f(s_L) + g(s_H)$ for all scales c′ more concave than c, so long as f and g are positive and increasing functions of s_L and s_H, respectively. In this case, inequality would be greater in y than in x regardless of whether f and g were chosen to emphasize inequality above the median, below the median, or if the weights were equal in this respect (as with the index s). It is not difficult to see why, in the case of health, it might be desirable to de-emphasize inequality above the median. Indeed, one may reasonably criticize health inequality rankings on the grounds that they 'reward' populations in which relatively fewer individuals are in excellent health. In some contexts, then, it may be desirable to choose the functions f and g so as to focus on inequality below the median. The next section demonstrates how measures of inequality that emphasize inequality below the median can, under certain conditions, be re-interpreted as measures of poverty, providing a useful extension of the results presented above.

3.3 Second order poverty comparisons

Consider a function $I'(s_L, s_H) = f(s_L) + g(s_H)$ that completely ignores inequality above the median, so that the $g(s_H)$ carries with it a value of 0 for all values of s_H. In this extreme case, the function $I'(s_L, s_H)$ is a positive monotonic transform of s_L, which is

---

[12]For example, a scale that exaggerated the distance between c_2 and c_3, or the distance between c_4 and c_5, might cause a re-ordering of x and y according to s_L or s_H.

equal to the median minus the average level of health for those below the median. So defined, s_L is simply the average health poverty gap, assuming the median has been selected as the health poverty threshold. Hence, $I(s_L, s_H)$ ranks distributions according to the health poverty gap, which is a measure of the depth of poverty, indicating as it does the average distance between the threshold and those who are below it.[13] It was noted above that if x S(c) y, then $I'(s_L(x;c), s_H(x;c)) < I'(s_L(y;c), s_H(y;c))$ for all scales c′ more concave than c. But if $I'(s_L(x;c), s_H(x;c)) < I'(s_L(y;c), s_H(y;c))$ for all reasonable scales, and if $g(\cdot)=0$, then $s_L(x;c) < s_L(y;c)$ for all reasonable scales. Apparently, when x S(c) y, x must have a lower average health poverty gap than y for all reasonable scalings of the SRHS data.

The spread index s_L may also impart poverty gap rankings that are robust to scale even when two distributions are not related by the second order S relation, since the ordering of the two distributions by the upper spread index s_H, which is critical to the establishment of a second order S relation, does not affect the poverty relationship. Take as an example distributions q and z with a common median m, where q first order dominates z. Since a greater proportion of z must have the highest level of health (e.g., excellent SRHS), q cannot also be more equal than distribution z. Yet if q first order dominates z, then q must also second order dominate z using scale c, which implies that the lower spread of q must be less than the lower spread for z, $s_L(q;c') < s_L(z;c')$, for all scales c′ at least as concave as c.[14] If c is a linear scale, then the average poverty gap is

---

[13]For further explanation, see Foster and Sen (1997).

[14]If q SD(c) z, then equation (25) holds. Note that (27) must also hold, since it is equal to (26) for categories $k \leq m-1$, where m is the median category, so that $s_L(q;c') <$

lower in distribution q than in distribution z as calculated using any reasonable scale. Thus, whenever a second order or second order S relationship exists between two distributions, an ordering of the lower spread index s_L may be established, and hence a poverty gap ranking can also be established that is robust to all reasonable scalings of the SRHS data.

Poverty gap rankings based on a second order relationship are also found to be robust to the choice of a poverty threshold. Let the poverty threshold be represented as category k+1. Note that equation (27), which applies whenever x SD(c) y or y S(c) x, requires overall (average) health in the portion of the distribution that lies below category k + 1 to be greater in x than in y, where k+1 is *any* category up to and including the median. When (27) applies, the difference between the c_{k+1} and the average level of health for those below category k+1 must be greatest in distribution y. This difference is just the poverty gap for a threshold of k+1. It is thus apparent that when (27) applies, the average poverty gap will be greater in y than in x so long as the poverty threshold is not above the median. Hence, when (27) applies, the poverty gap ranking of x and y is robust both to the use of alternative (reasonable) scales, and to the choice of alternative poverty thresholds, so long as the threshold is not above the median level of health.

4. Results

Age- and sex-adjusted SRHS data for the 25 largest states from the 1994 *NHIS State Data Files* are used below to search for second order and second order S

$s_L(z;c')$, for all scales c' at least as concave as c.

relationships between states. The main objective is to generate new distributional relationships between states and to see whether these new relationships strengthen, confirm, or cast doubt upon the patterns identified in chapter III, which are obtained using the first order and first order S dominance relations. The reference scale used to generate orderings using the SD(c) and S(c) relations is the linear c=[1,2,3,4,5] scale so that results apply to all reasonable scalings of the data. These results are then used to identify robust health poverty rankings that take into account the depth of poverty.

Table 7 summarizes the results of a search for state-to-state second order relationships. A '1' in a cell indicates that the column state second order dominates the row state, and thus has more per-capita health than the row state for any linear or concave scaling of the SRHS data. The total number of second order relations in the 25 state sample is 257, compared to the 199 state-to-state relationships generated using first order dominance as in chapter III. These 58 new comparisons of overall health facilitated by the restricted scale approach introduced above represent a 29% increase over the scale independent approach introduced in chapter III. Note that the pattern in Table 7 strongly indicates that southern states have less health than states in other regions of the country. Among the new comparisons of overall health facilitated by the use of second order dominance, dominant states are disproportionately located in non-southern regions, while dominated states are disproportionately located in the South.[15] Hence, the new comparisons facilitated by the restricted scale approach strengthen the result

[15]Over four-fifths of the 58 new relationships place non-southern states in the dominant position, though these states comprise only three-fifths of the states included in the study; and in 57% of the newly-related state-pairs southern states have less overall health, though only 40% of states are in that region.

Table 7. Second order comparisons of overall SRHS among 25 largest states. A '1' in a cell indicates that the column state has more overall health than the row state according to the SD(c) relation using a linear reference scale (data source: 1994 *NHIS State Data Files*).

	Northeast				Midwest							West				South										
	MA	NY	NJ	PA	OH	IN	IL	MI	WI	MN	MO	CA	WA	CO	AZ	MD	VA	NC	GA	FL	AL	TN	LA	TX	OK	Totals
Northeast																										
MA		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NY	1		0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	4
NJ	1	1		0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	7
PA	1	1	0		1	0	1	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	9
Midwest																										
OH	1	1	0	0		0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	6
IN	1	1	1	1	1		1	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	11
IL	1	1	0	0	0	0		0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	5
MI	1	1	1	1	1	1	1		1	1	1	0	1	1	0	1	1	0	0	1	0	0	0	0	1	16
WI	1	0	0	0	0	0	0	0		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
MN	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MO	1	1	1	1	1	1	1	0	1	1		0	1	1	0	1	1	0	0	0	0	0	0	0	0	13
West																										
CA	1	1	1	1	1	1	1	0	1	1	1		1	1	0	1	1	0	0	0	0	0	0	0	0	14
WA	1	0	0	0	0	0	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0	0	2
CO	1	0	0	0	0	0	0	0	0	1	0	0	0		0	0	0	0	0	0	0	0	0	0	0	2
AZ	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	0	1	1	0	0	0	1	1	20
South																										
MD	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	2
VA	1	1	0	1	1	0	1	0	1	1	0	0	1	1	0	1		0	0	0	0	0	0	0	0	10
NC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	0	1	1	1	1	23
GA	1	1	1	1	1	0	1	0	1	1	0	0	1	1	0	1	1	0		0	0	0	0	0	1	13
FL	1	1	1	1	1	0	1	0	1	1	0	0	1	1	0	1	1	0	0		0	0	0	0	0	12
AL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	24
TN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0		0	1	1	20
LA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0		1	1	20
TX	1	1	1	1	1	0	1	0	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0		0	12
OK	1	1	0	1	1	0	1	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0		10
Totals	23	18	12	14	16	8	15	5	17	23	7	5	19	15	2	19	11	1	5	6	0	2	2	5	7	257

obtained using the scale independent method, which indicated that southern states have the least overall health.

Table 8 summarizes results of a comprehensive search for S(c) inequality relationships in the 25 state sample. Using a linear reference scale, 65 second order S relationships were found between states in the sample, representing an 18% increase over the 55 first order S relationships identified in chapter III. Again, of the new comparisons facilitated by the second order S relation, dominant (less equal) states were disproportionately located in the south, while dominated (more equal) states were disproportionately located in non-southern regions.[16] Hence, excluding unreasonable non-concave scalings of the data strengthens the earlier results that used first order S dominance without any scale restrictions: southern states have a more unequal distribution of health than states in other regions of the country – especially when compared to states in the northeast and midwest.

Results from the search for poverty gap orderings among the 25 state sample are presented in Table 9. A '1' in a cell indicates that the column state has a lower average poverty gap than the row state for all reasonable scalings and for all poverty thresholds less than or equal to the median category of very good health. Poverty gap orderings were obtained in 90% of the 300 possible state-pair comparisons. As with the poverty rate orderings established in chapter III, these comparisons overwhelmingly favor non-

[16]Four-fifths of the 10 new state-to-state relationships place non-southern states in the more equal position, though these states only comprise three-fifths of the included states; and four-fifths of the 10 new state-to-state relationships place southern states in the less equal position, though these states comprise only two-fifths of the number of states in the sample.

Table 8. Second order S inequality comparisons in SRHS among the 25 largest states. A '1' in a cell indicates that the column state has more equal health than the row state according to the S(c) relation using a linear reference scale (data source: 1994 *NHIS State Data Files*).

	Northeast				Midwest							West				South										
	MA	NY	NJ	PA	OH	IN	IL	MI	WI	MN	MO	CA	WA	CO	AZ	MD	VA	NC	GA	FL	AL	TN	LA	TX	OK	Totals
Northeast																										
MA		0	0	0	0	0	0	0	1	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NY	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NJ	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PA	0	1	0		0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Mldwest																										
OH	0	1	0	0		0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
IN	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IL	0	1	0	0	0	0		0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
MI	0	0	1	0	0	1	0		0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	4
WI	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MN	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MO	0	0	0	0	0	1	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
West																										
CA	0	0	0	0	0	1	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0	0	0	2
WA	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0
CO	0	0	0	0	0	0	0	0	0	1	0	0	1		0	1	0	0	0	0	0	0	0	0	0	3
AZ	0	0	1	0	0	1	0	1	1	0	1	1	0	0		0	0	0	0	0	0	0	0	0	0	7
South																										
MD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
VA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0
NC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0
GA	0	1	1	1	1	1	1	0	1	0	1	0	0	0	0	0	1	0		1	0	0	0	0	1	11
FL	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0		0	0	0	0	0	5
AL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0
TN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0
LA	0	1	1	0	0	1	1	1	1	0	1	0	0	0	0	0	1	0	0	1	0	0		0	1	11
TX	0	1	1	1	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0		0	8
OK	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0		5
Totals	0	8	7	2	1	9	4	2	9	2	6	2	1	0	0	1	7	0	0	2	0	0	0	0	2	65

Table 9. Robust poverty gap comparisons in SRHS among the 25 largest states. A '1' in a cell indicates that the column state has a lower health poverty rate and a lower health poverty gap than the row state for all reasonable scales and for all poverty thresholds of very good health or less (data source: *1994 NHIS State Data Files*).

	Northeast				Midwest							West				South										
	MA	NY	NJ	PA	OH	IN	IL	MI	WI	MN	MO	CA	WA	CO	AZ	MD	VA	NC	GA	FL	AL	TN	LA	TX	OK	Totals
Northeast																										
MA		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NY	1		0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	4
NJ	1	1		0	1	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	7
PA	1	1	0		1	0	1	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	9
Midwest																										
OH	1	1	0	0		0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	6
IN	1	1	1	1	1		1	0	1	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	11
IL	1	1	0	0	0	0		0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	6
MI	1	1	1	1	1	1	1		1	1	1	0	1	1	0	1	1	0	0	1	0	0	0	0	1	16
WI	1	0	0	0	0	0	0	0		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
MN	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MO	1	1	1	1	1	1	1	0	1	1		0	1	1	0	1	1	0	0	0	0	0	0	0	0	13
West																										
CA	1	1	1	1	1	1	1	0	1	1	1		1	1	0	1	1	0	0	0	0	0	0	0	0	14
WA	1	0	0	0	0	0	0	0	0	1	0	0		0	0	0	0	0	0	0	0	0	0	0	0	2
CO	1	0	0	0	0	0	0	0	0	1	0	0	1		0	1	0	0	0	0	0	0	0	0	0	4
AZ	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	0	1	1	0	0	0	1	1	20
South																										
MD	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0		0	0	0	0	0	0	0	0	0	2
VA	1	1	0	1	1	0	1	0	1	1	0	0	1	1	0	1		0	0	0	0	0	0	0	0	10
NC	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	0	1	1	1	1	23
GA	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	1	0		1	0	0	0	0	1	16
FL	1	1	1	1	1	1	1	0	1	1	0	0	1	1	0	1	1	0	0		0	0	0	0	0	13
AL	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	24
TN	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0		0	1	1	20
LA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0		1	1	20
TX	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	0	0	0	0		0	14
OK	1	1	1	1	1	1	1	0	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0		13
Totals	23	18	13	14	16	12	15	5	18	24	9	5	20	15	2	20	12	1	5	7	0	2	2	5	7	270

southern states.[17] Taken together, these results indicate that both the prevalence and the depth of health poverty are greatest in southern states and that this finding is not subject to reasonable alternative choices of scale or poverty threshold.

5.Conclusion

This paper presents a new methodology for ranking qualitative distributions that takes advantage of scale restrictions that may be applied when using certain types of data. For SRHS data, it is argued that any reasonable numeric scaling of the data should be concave. Adapting a common result from the income measurement literature, a methodology for ranking distributions is introduced that depends upon the establishment of a second order dominant relationship between (categorical ranges of) two distributions. It is shown that when one distribution second order dominates another using the linear scale, it has a higher overall -- or mean -- level of health using any concave scale. In addition, a measure of dispersion in health is introduced, second order S dominance, that also holds over all reasonable scales. Finally, the relationship between second order dominance, second order S dominance, and the poverty gap is described, leading to the establishment of poverty rankings that are robust to both reasonable

---

[17]The poverty rate is also referred to as the headcount ratio, and is equal to the proportion of a given population that lies at or below the poverty level. Note that since poverty gap orderings exist whenever equation (27) applies, and (27) applies whenever x first order dominates y up to the median, the FD and S relations may also be used to establish a poverty gap ordering between x and y. In particular, (27) applies if either x FD y or y S x. Hence, the poverty *rate* orderings found in chapter III, which exist when an FD or S ordering exists, also impart poverty *gap* orderings. Nevertheless, since (27) ignores the lower mean orderings associated with convex scales, while the FD and S relations do not, (27) may be able to establish a poverty gap ordering where the FD and S relations do not apply.

alternative scales and reasonable alternative poverty thresholds.

These new distributional measures are used to test the representativeness of the results obtained in chapter III, which suggest that health levels are lowest, and health inequality and health poverty rates highest, in southern states. The new methodology yields roughly a 20% increase in the number of state pairs in which an inequality ranking can be established, and a 30% increase in the number of comparisons of overall health. These new comparisons strengthen the earlier findings that health levels are lowest and health inequality highest in southern states. Moreover, it is shown that not only are health poverty *rates* highest in southern states, but the *depth* of health poverty is also greatest in the south, independent of any reasonable choice of scale or poverty threshold.

Despite the significant increase in measurement power achieved with the scale-restricted approach presented in this chapter, some may consider the specific scale restrictions employed in this chapter to be too strong, too weak, or somehow inaccurate. The method is, however, flexible in that the reference scale used to establish second order dominance need not be linear. By choosing a different reference scale, one could further restrict (or expand) the set of scales over which distributional comparisons must apply, potentially increasing (or decreasing) the number of comparisons that can be made. Nevertheless, such flexibility may yet prove inadequate in certain contexts, underscoring the need for further research into methods of this type.

CHAPTER V

EXPLAINING BETWEEN-STATE VARIATION IN SELF-REPORTED HEALTH STATUS

1. Introduction

The preceding chapters use qualitative health data to illuminate key regional patterns in the distribution of health across states. This chapter demonstrates how aggregated qualitative health data can be used to measure and explain differences in overall health across states. The motivation for this analysis is, in part, to provide information that could be used by federal policy makers to better target public health dollars such as those disseminated to states to help finance the Medicaid program. The central policy question addressed in this chapter is whether Medicaid eligibility rules are designed to target federal Medicaid funds toward states with the greatest concentration of health needs. An interesting follow-up question would be whether federal Medicaid funding is concentrated among states with the greatest health needs. Yet devising an informative analysis of federal spending patterns is problematic. Since states have a great deal of flexibility in designing, administering and funding their Medicaid programs, the federal government does not directly control the flow of funds *within* states, nor the resulting flow of funds *to* states. As a result, this chapter focuses on program design issues that affect the flow of federal funds. Specifically, the chapter examines whether the federal government's eligibility guidelines *enable* funds to flow toward states with the greatest health needs.

The key policy tools available to the federal government in designing the

Medicaid program and determining the flow of funds to states are: (i) determining the grant formula that provides aid to states for Medicaid expenditures; (ii) defining Medicaid benefits, e.g., deciding which types of individuals are eligible and which are not; (iii) designating program components as either mandatory or optional for each state; and (iv) establishing guidelines for program delivery and administration, e.g., determining which benefits may be provided under capitation. Reports prepared by the Urban Institute (1993) and the General Accounting Office (1991, 1993) address the distribution of federal Medicaid funds across states, focusing in particular on various effects of the federal grant formula. The grant formula matches state Medicaid expenses at the rate of 1 - .45 * (state per-capita income / U.S. per-capita income)^2, with an imposed floor of 50% and a ceiling of 83%. Matching rates thus increase at an increasing rate as relative state per-capita income declines. Since income and health are generally related at both the individual and aggregate level, it might be assumed that the matching rate increases along with state health needs, though not necessarily at the same rate. Nevertheless, the Urban Institute report suggests that the formula might better adjust for state health needs if it were to take into account "the distribution of poverty, the age distribution, and the prevalence of persons with disabilities and/or specific morbidities."[1]

The Urban Institute and GAO reports focus on one tool of federal policy, the grant formula, in evaluating the distribution of federal aid. This chapter evaluates another tool of federal Medicaid policy, the definition of Medicaid benefits.

---

[1]Blumberg, Holahan, and Moon (1993), p. 47.

Specifically, this chapter addresses the question of whether the federal government has designed benefits that *enable* the flow of funds to the neediest states. Given that federal funding varies according to the level of benefits provided in each state, this question can be posed directly in terms of the benefit structure: are Medicaid benefits designed to provide an opportunity for states to meet the health needs of their populations? Note that this analysis focuses on benefit rules and does not address the impact that other tools of federal policy have on states' ability to help the neediest individuals, such as the designation of benefits as mandatory rather than optional or the design of the federal grant formula.

One may reasonably question whether the federal government has adopted the objective which serves as the standard of evaluation in this analysis, namely the establishment of policies that reduce (or redress) differences in health needs across states. Unfortunately, Congressional intent in designing the Medicaid program is difficult to ascertain. The Urban Institute summarized the dilemma in its 1993 report on the Medicaid funding formula:

> Medicaid program objectives have suffered from a lack of clarity and a diversity of perspectives. Historically speaking, the legislative intent was ambiguous. And policy analysts, states, and the federal government continue to hold divergent views of the role of Medicaid in the health care/welfare system. Without a uniform understanding of its objectives, it should not, perhaps, be surprising that Medicaid has failed to fulfill the expectations of many different interests.

Despite this ambiguity, the General Accounting Office (1991, 1993) emphasizes the federal goal of reducing program benefit disparities across states, a move that would target funds towards the neediest states, assuming that benefits are designed to cover individuals with the greatest health needs.

There is at least some evidence that the federal government sought to reduce benefit disparities during the program's formative years. In 1955, a Library of Congress analyst describing changes to Medicaid's predecessor, Medical Assistance for the Aged (MAA), which is also known as the "Kerr-Mills" program, noted that "The Federal standards named in the bill are designed to encourage more uniformity in State programs"[2] And, speaking to the House in favor of increases in funding for the MAA program on July 31, 1958, Congressman Wilbur Mills noted that the changes:

> should put the additional funds where they are most needed and should serve the national interest by reducing differences in the size of payments to needy people among States and to help to assure that no needy person will go without the bare necessities of life.

Even if these views could be considered representative of federal goals, it would be difficult to discern Congressional intent without a specific definition of need. Though Congressional intent remains clouded in this respect, federally-defined benefits may nevertheless be examined to determine the extent to which the design targets states with

[2]Helen E. Livingston (1956), p. 73.

the greatest *health* needs.

At least two recent analyses address health differences across states. Kennedy et al (1998) use cross-sectional Self-Reported Health Status (SRHS) data from the Behavioral Risk Factor Surveillance System to demonstrate the effects of state income inequality on individual health. Ruhm (1996) uses time series data to evaluate the effects of recessions on state mortality rates. The following sections of this chapter address the specific question of whether the primary determinants of Medicaid eligibility – established by Congress – coincide with the demographic characteristics that explain differences in health needs across states. These policy questions affect the design of the econometric model employed and distinguish this analysis from previous efforts to explain differences in health across states. Another distinguishing characteristic of this analysis is its use of SRHS data to represent aggregate health levels. The remainder of this chapter proceeds as follows. Section 2 presents a methodology for explaining differences in overall health across states using qualitative health data. Section 3 specifies the model used to explain between-state differences in health. Section 4 presents the results of this analysis and demonstrates the significant explanatory power of a major demographic characteristic – educational attainment – that does not directly affect Medicaid eligibility. Section 5 summarizes the results of this chapter and discusses the policy implications of the results.

2. Identifying health needs among the states

There are many ways in which one might define and measure the health needs of a state, e.g., mortality rates, specific disease prevalence, age profiles, the depth or

prevalence of poverty, and so forth. The preceding chapters advocate the use of SRHS as an indicator of general health status, and introduce methods for comparing overall levels of health in different populations. The results from distributional comparisons employing these methods show a pattern of regional disparities in health in which southern states have less overall health and greater health poverty than states in other regions. Current Medicaid financing mechanisms favor poorer states, which are disproportionately located in the south. Nevertheless, even if the financing mechanism successfully equalizes the financial capacity of each state to provide Medicaid benefits, states' ability to meet the health needs of their population will differ if Medicaid benefits are not targeted at the greatest health needs in each state. Medicaid eligibility is determined primarily by one's income, age, and family structure, e.g., whether a wage-earning spouse is present. Special eligibility categories also exist for pregnant women.[3] Do Medicaid benefits successfully account for differences in health needs across states? To help answer this question, it would be helpful to know if state health needs differ significantly by identifiable characteristics *other* than those the Medicaid program takes into account in its coverage rules. In particular, are there factors other than income, age, family structure and pregnancy that significantly affect the overall level of health in a state's population?

How might differences in health across states, as measured by SRHS data, be

---

[3]In addition, specific eligibility provisions exist to accommodate those receiving institutionalized long term care. A significant proportion of Medicaid expenditures go towards institutionalized long term care for the elderly. This paper does not address the needs of this population, nor the effect of this population on state health burdens, since the data set employed samples only the non-institutionalized population.

characterized so that standard statistical techniques may be used to help explain the differences? The approach taken in this paper is to use as the dependent variable the cumulative percentages of SRHS in each state at the various levels of SRHS. These percentages amount to headcount ratios, or poverty rates, and can alternatively be thought of as the probability that an individual in a particular state will be in poor health, where poor health is defined by the chosen threshold. To avoid the subjective choice of a poverty threshold, this analysis compares results for each of the four possible thresholds: poor, fair, good, and very good SRHS. This approach may be used to establish the robustness of results to the choice of health poverty threshold, and may be contrasted with Kennedy et al's (1998) analysis of health poverty at the individual level, whose results are contingent upon their use of a single threshold of fair SRHS. If a characteristic were found to explain changes in health using each of the four dependent variables, then the characteristic must have a first order effect on SRHS, which is to say that changes in the characteristic result in an upward or downward shift in the distribution of SRHS, and have a concomitant effect on mean SRHS regardless of how one chooses to scale the data. For those characteristics that are not found to have a first order effect on health, the use of multiple models using four different dependent variables enables the identification of health effects that vary over the range of health status.

3. Specification

The explanatory model used in this analysis is intended to pick up the effect on state health levels of demographic characteristics reflected in Medicaid eligibility rules,

and to identify characteristics affecting state health levels that have not been directly incorporated into these rules. Due to the small number of observations in the data set used to estimate it, the model is specified in as parsimonious a fashion as seems reasonable. The standard model estimated is of the following form:

$$h_{it} = \alpha + X_{it}\beta + Y_{it}\Gamma + T\eta + U_{it}\gamma + \epsilon_{it}, \tag{29}$$

where h_{it} is the overall level of health in state i at time t, the vector X includes state demographic characteristics that are reflected in Medicaid eligibility criteria, Y includes demographic characteristics that do not affect Medicaid eligibility, T is a vector of t-1 year dummies, and U is the unemployment rate.

Overall health in the state, h, is represented by four variables equal to the percentages of a state's population with SRHS of poor, fair or less, good or less, and very good or less. The values for these variables in 1994 are presented in Table 10. These cumulative percentages can most easily be understood as health poverty rates, where the poverty line is set equal to, respectively, poor, fair, good, and very good SRHS. It may be inappropriate to characterize those with very good SRHS as being in health poverty, especially since very good SRHS is the universal median level of health in the data set used in this analysis. Since the cumulative percentage of the population with SRHS of very good health imparts the same information as the percentage of the population with excellent SRHS, the dependent variable can in this case be characterized as explaining the rate of health excellence in states, though the reported sign of the coefficients would be reversed.

Table 10. Cumulative percentages at at each level of self reported health status

Northeast	Poor	Fair or less	Good or less	Very Good or less
Massachusetts	1.78%	7.30%	28.44%	54.79%
New York	2.28%	9.11%	30.26%	62.83%
New Jersey	2.52%	9.72%	35.01%	64.19%
Pennsylvania	2.65%	9.93%	33.05%	61.51%
Midwest				
Ohio	2.44%	9.64%	33.02%	61.18%
Indiana	2.76%	10.11%	35.46%	65.53%
Illinois	2.62%	9.11%	31.17%	61.69%
Michigan	3.55%	11.98%	35.43%	63.84%
Wisconsin	1.85%	8.98%	31.07%	63.61%
Minnesota	1.49%	7.24%	25.95%	57.81%
Missouri	3.07%	11.10%	35.41%	65.00%
West				
California	3.23%	11.11%	37.30%	64.73%
Washington	2.19%	8.90%	27.55%	59.46%
Colorado	2.35%	10.01%	28.25%	55.00%
Arizona	4.00%	13.28%	37.95%	63.27%
South				
Maryland	2.24%	7.97%	27.66%	59.10%
Virginia	3.02%	10.17%	34.01%	64.07%
North Carolina	4.89%	13.83%	41.58%	66.78%
Georgia	3.93%	12.93%	34.08%	60.58%
Florida	3.40%	10.35%	36.83%	61.96%
Alabama	5.61%	17.94%	44.01%	69.45%
Tennessee	4.51%	12.44%	39.24%	66.05%
Louisiana	4.78%	13.34%	37.09%	61.59%
Texas	3.14%	12.22%	36.18%	61.43%
Oklahoma	3.48%	11.41%	33.56%	62.27%

Characteristics included in X are intended to provide a succinct proxy for Medicaid eligibility criteria for the non-institutionalized population and include variables representing income, family structure, and pregnancy status. The data are age- and sex-adjusted, as described in Chapter III, so age and sex variables are not included in the model. Results from numerous studies indicate that health increases at a decreasing rate with income, i.e., there is a non-linear *concave* relationship between income and health[4]. Hence, a squared per-capita income variable was included along with per-capita income. The variable representing family structure is the percentage of children living in one-parent families, since these families form the basis for traditional welfare-related Medicaid eligibility. A variable representing the health burden of pregnancy was included to accommodate the significant expansions of Medicaid coverage to poor and near-poor pregnant and postpartum women in the early 1990s. The variable selected to represent the health burden of pregnancies in a state is the teen pregnancy rate, since health concerns associated with pregnancy and childbirth are expected to be greatest in this population of young mothers.

Demographic characteristics included in Y do not directly affect Medicaid eligibility and include variables representing educational attainment and racial composition. Educational levels are characterized by three variables representing the attainment of three different levels of education for the state: the percentage of state residents with less than a high-school degree, the percentage attending college, and the percentage that attained at least a college degree. The omitted category was the

---

[4]See, for example, McDonough, Duncan, Williams, and House (1997), Backlund, Sorlie and Johnson (1996), and House, Kessler and Herzog (1990).

percentage obtaining a high school diploma. Racial composition is represented by two variables: the percentage of the population that is black and the percentage of the population comprised of other non-whites. The omitted category is the percentage of the population that is white.

The unemployment rate U is included due to Ruhm (1996), who found a *positive* and significant contemporaneous effect of state unemployment levels on state mortality rates. Unemployment rates for each state are for the month of June and are not seasonally adjusted. The t-1 time dummies in T, where 1990 is the excluded year, account for any remaining shifts in the distribution of SRHS over time.

4. Results

The explanatory variables described above were used in eight regressions estimated using ordinary least squares. The data set used in this analysis consists of a pooled cross-section of annual state-level data over the 1990-1994 period. The data are from the *NHIS-SDF*, which is described in more detail in chapter III. The *NHIS* samples only the non-institutionalized population, providing a natural control for the significant differences in state health needs that may arise due to the distribution of institutionalized individuals. The data include the same 25-state selection used in previous chapters,[5] which leaves 125 observations over five years.

The first four models, presented in columns (a)-(d) of Table 11, show the average effects of each variable on the cumulative percentage of a state's population with poor

[5]See footnote 28 in chapter III for further explanation.

Table 2. Effects of state characteristics on rates of health poverty (OLS)[1]

	I Threshold: Poor SRHS		II Threshold: Fair SRHS		III Threshold: Good SRHS		IV Threshold: Very Good SRHS	
	b	*t*	b	*t*	b	*t*	b	*t*
CONSTANT	0.034	1.36	0.169	2.70	0.139	1.21	0.489	4.33
PER CAPITA INCOME ($000s)	(0.006)	3.38	(0.021)	4.46	(0.019)	2.29	(0.007)	0.86
PER CAPITA INCOME^2	0.000	3.20	0.000	4.64	0.000	2.34	0.000	1.26
BLACK	0.211	1.90	0.033	1.16	(0.078)	1.50	(0.096)	1.89
OTHER NON-WHITE	(0.014)	0.59	(0.066)	1.09	(0.040)	0.36	0.010	0.09
NO H.S. DEGREE	0.092	2.52	0.357	3.89	0.938	5.57	0.723	4.37
SOME COLLEGE	0.055	1.24	0.263	2.37	0.428	2.10	0.355	1.77
COLLEGE GRADUATE	0.021	0.55	(0.049)	0.50	0.107	0.60	(0.612)	3.49
TEEN PREGNANCY RATE	0.000	2.32	0.000	0.73	0.000	0.04	(0.000)	0.56
KIDS WITH ONE PARENT	0.000	1.26	0.000	0.06	0.002	1.85	0.001	0.84
UNEMPLOYMENT RATE	(0.011)	0.22	0.021	0.17	0.034	0.15	0.707	0.31
Year dummies								
1991	0.001	0.70	0.008	1.85	0.014	1.89	0.009	1.24
1992	0.003	1.65	0.013	2.79	0.020	2.21	0.010	1.13
1993	0.003	1.85	0.018	4.05	0.034	4.21	0.016	1.95
1994	0.006	3.64	0.017	3.81	0.027	3.38	0.010	1.21
R^2	0.782		0.765		0.757		0.660	

Notes:

1. Dependent variables are cumulative percentages of each state's population with health less than or equal to the threshold.

(column a), fair (column b), good (column c), and very good (column d) health. All of the variables in equation (29) are included in these models. Coefficients reflect the estimated percentage point change in health poverty (or health excellence) rates due to a unit increase in the explanatory variable.

The income and income-squared variables are significant in each of the regressions where the threshold is less than the median level of health, indicating that income has a negative and diminishing effect on health poverty rates, and no significant effect on rates of health excellence. These results suggest that higher levels of state income are associated with lower levels of within-state dispersion in health, though the association would vary according to the measure of dispersion used and the SRHS scale employed.[6]

At least one educational attainment variable is significant in all four specifications, and the signs on the significant education variables are generally as expected. The percentage of residents without a high school diploma increases health poverty rates and decreases rates of health excellence. Increases in the proportion of a state's population with a college degree increase rates of health excellence. Curiously, though, increases in the percentage of the population with some college experience tend to *increase* health poverty rates.

Racial composition, as measured by the two variables representing the percentage

---

[6]The coefficient on income in model (d) is negative, indicating an (insignificant) spread in health above the median as income increases. As a result, the robust inequality measures discussed in chapters III and IV would not register decreases in dispersion due to income. Whether or not the effects of income above the median would offset the equalizing effects of income below the median is thus dependent upon the choice of index and the scaling of the SRHS data.

of the population that is not white, is not significant in any of the models. Neither do the two race variables have a jointly significant effect on explanatory power in the models (results not shown). The teen pregnancy rate was only significant in model (a), which explains extreme health poverty, while the percentage of children in one-parent households was not significant in any of the four models. Finally, note that the unemployment variable has an insignificant effect on population health in each of the models, which may suggest that Ruhm's (1996) results are dependent upon the choice of health indicator.[7]

Just one variable, the percentage of the state's population without a high school degree, is significant in all four of the models presented above. This implies that only this variable has an independent first order effect on levels of health in a state. Hence, none of the other variables in the model could have a significant effect on mean state health levels independent of the other variables and independent of the SRHS scale used to calculate mean health.[8]

The principle question of this analysis is whether the Medicaid program adequately takes into account differences in health needs across states in the design of its benefits. One of the variables found to be a significant explanator of state SRHS levels – educational attainment – is not taken directly into account in Medicaid's eligibility rules.

---

[7]Ruhm's health indicator is mortality. His results are more robust in that they draw upon a much longer time series. Moreover, since unemployment and income affect the fiscal capacity of states to provide benefits, one should not draw from the results presented in this paper the conclusion that Medicaid funding should be insensitive to the specific economic conditions of each state.

[8]This inference was confirmed in a series of regressions (not reported) using means generated by numerically scaled SRHS data.

The other demographic characteristic that Medicaid ignores, race, is not a significant explanator of health differences across states. Together, the education and race variables are found to significantly increase the explanatory power of each model.[9] As an example, note that the R^2 for model (a) is .78 when the education and race variables are included, but falls to .72 when they are removed, while the R^2 for model (d) falls from .66 to .45 when these variables are removed (results not shown). Apparently, the explanatory power of the two variables, especially the educational variables, rises as the poverty threshold in the dependent variable rises.

5. Discussion

This paper presents an analysis of between-state differences in health. The analysis may be used to evaluate the success of the federal government in designing benefits that enable states to meet the health needs of their populations. The paper addresses the specific question of whether there is a close correspondence between the major determinants of differences in health across states and the primary characteristics affecting Medicaid eligibility. The paper suggests the use of SRHS data as an indicator of state health levels in a series of models designed to identify results that have a first order effect on state health levels. The results indicate that only one of the variables included in the analysis -- the percentage of the state's population without a high school degree -- has a significant and independent first order effect on state health levels.

---

[9]Results not shown. An F-test of the null hypothesis that the true model does not include these variables resulted in an F values ranging from 6.87 when the poverty threshold for the dependent variable is poor SRHS, to 14.32 when the threshold is very good SRHS, compared to a significance level ($p<.01$) of about 3.2.

Hence, only this variable has an effect on mean health that is independent of the numeric SRHS scale used. The mix of influential characteristics changes as one moves up the health spectrum. For example, the income variables are highly significant explanators of the variation across states in health poverty rates with thresholds of poor or fair SRHS, while only education variables are significant explanators of variation in rates of health excellence.

One demographic characteristic that has a significant effect on various rates of health poverty and health excellence, educational attainment, merits further attention. Each of the other variables included in the analysis are reflected by some component of Medicaid eligibility rules. By providing benefits to those in 'needy' demographic categories defined according to age, income, family structure, and pregnancy status, the Medicaid program has already begun to target needy populations in various states. The more needy individuals a state has -- as defined by these characteristics -- the more potential Medicaid recipients it will have, and thus the more federal aid (in the form of matching funds) will be available to the state to help meet the health needs associated with these categories of individuals.

While the Medicaid program does not take educational attainment specifically into account in its eligibility criteria, this analysis finds, as numerous other studies find, that education has a significant and *independent* effect on health.[10] Economic theory provides a firm basis for understanding this result. Grossman's (1972) model of the demand for health suggests an individual health production process that may vary in

---

[10]See, for example, Wagstaff (1986a), van Vliet and van Praag (1987), Panis and Lillard (1994), and Hurd and McGarry (1995).

productivity according to levels of education. Various explanations have been offered as to how increases in education may make individuals more efficient producers of health. For example, better educated individuals may better understand medical instructions or have greater access to information.[11]

It may be difficult or undesirable to further refine Medicaid eligibility rules to directly account for between-state differences in health caused by variations in schooling. Instead, federal policy makers could use other policy tools to address interstate inequities in health needs, e.g., public health information campaigns. Though motivations may vary, such informational campaigns are indeed commonplace. Nevertheless, this study and others indicate that education is one of the *principle* determinants of health differences across states. Given the importance of education in explaining health, policy makers may want to re-direct public health funds towards informational services, even if these funds must come at the expense of health services.[12]

---

[11]Panis and Lillard (1994).

[12]See Wagstaff (1986b) for a related discussion.

APPENDIX A

CHAPTER III PROOFS

Theorem III.1 Let x and y be any two distributions. Then x FD y if and only if $\mu(x;c) \geq \mu(y;c)$ for every scale c.

Proof Let x and y be any two distributions. A simple calculation yields

$$\mu(x;c) = c_1 + \sum_{j=1}^{n-1}\left(c_{j+1} - c_j\right)\left(1 - X_j\right), \tag{30}$$

and so

$$\mu(x;c) - \mu(y;c) = \sum_{j=1}^{n-1}\left(c_{j+1} - c_j\right)\left(Y_j - X_j\right). \tag{31}$$

Now suppose that x FD y. Then, by definition, $Y_k - X_k \geq 0$ for all k=1,...,n-1, and hence $\mu(x;c) - \mu(y;c) \geq 0$ for any scale c. Conversely, if x FD y does *not* hold, then there must be some k=1,...,n-1 for which $Y_k - X_k < 0$. Clearly there must exist a scale $c'=[c'_1,...,c'_n]$ for which the inter-category gap $c'_{k+1} - c'_k > 0$ is so large relative to the other inter-category gaps as to render the entire sum $\sum_{j=1}^{n-1}\left(c'_{j+1} - c'_j\right)\left(Y_j - X_j\right)$ negative. Therefore, $\mu(x;c') - \mu(y;c') < 0$, which ensures that the condition "$\mu(x;c) \geq \mu(y;c)$ for every scale c" fails to hold.

Theorem III.2 Let x and y be two distributions having the same median category. Then x

S y if and only if the S-curve of x is no lower than the S-curve of y.

Proof Let x and y be any two distributions with the same median category m. Suppose that x S y holds, so that $X_j \geq Y_j$ for all $j < m$ and $X_j \leq Y_j$ for all $j \geq m$. For ease of notation, denote $X_0=Y_0=0$. Select any population share p between 0 and 1 that is not equal to X_j or Y_j for any j=0,1,...,n. Then there must exist j and j′ such that $X_{j-1} < p < X_j$ and $Y_{j'-1} < p < Y_{j'}$, and so by definition, the S-curve value at p is $|m-j|$ for x and $|m-j'|$ for y. Now if $j < m$, then since $X_j \geq Y_j$ we must have $j' > j$, which ensures that $|m-j| \geq |m-j'|$. Similarly, if $j > m$, then $X_j \leq Y_j$ we must have $j' < j$, which leads once again to $|m-j| \geq |m-j'|$. The remaining case where $j = m$ must entail $Y_{j'-1} \leq X_{j-1} < X_j \leq Y_j$, and $j' = j$, and so $|m-j| = |m-j'| = 0$. Therefore, at all such p, the S-curve of x is not below the S-curve of y. This relationship clearly extends to the remaining values of p by definition of the S-curve.

To show the converse, suppose that x and y do not satisfy x S y. Then for some $j < m$ we must have $X_j < Y_j$ or for some $j \geq m$ we must have $X_j > Y_j$. Pick p between X_j and Y_j. If $j < m$, then at p the value of the S-curve of x must be less than $|m-j|$, while the value of the S-curve of y must be at least $|m-j|$, ensuring that the S-curve of y is above the S-curve of x at p. Similarly, if $j \geq m$, then at p the value of the S-curve of y must be greater than $|m-j|$ while the value of the S-curve of x must be no greater than $|m-j|$, once again implying that at p, the S-curve of y is above

the S-curve of x. In either case, the S-curve of y is above the S-curve of x at some p, and so the result is verified.

Theorems III.3 and III.4 Let x and y be any two distributions that have the same median category. Then the following are equivalent:

(i) x S y.

(ii) $s(x;c) \geq s(y;c)$ for every scale c.

(iii) $s_L(x;c) \geq s_L(y;c)$ and $s_H(x;c) \geq s_H(y;c)$ for every scale c.

Proof Suppose that x and y have the same median category m. A simple calculation yields

$$\mu_L(x;c) = c_1 + \sum_{j=1}^{m-1}\left(c_{j+1} - c_j\right)\left(1 - 2X_j\right) \text{ and} \tag{32}$$

$$\mu_H(x;c) = c_m + \sum_{j=m}^{n-1}\left(c_{j+1} - c_j\right)\left(1 - 2X_j\right) \tag{33}$$

and so

$$\mu_L(x;c) - \mu_L(y;c) = 2\sum_{j=1}^{m-1}\left(c_{j+1} - c_j\right)\left(Y_j - X_j\right) \text{ and} \tag{34}$$

$$\mu_H(x;c) - \mu_H(y;c) = 2\sum_{j=m}^{n-1}\left(c_{j+1} - c_j\right)\left(Y_j - X_j\right). \tag{35}$$

(i)↔(iii): Suppose that x S y. Then, by definition, $X_j - Y_j \geq 0$ for all j=1,...,m-1 and $X_j - Y_j \leq 0$ for all j=m,...,n-1, and hence $\mu_L(x;c) - \mu_L(y;c) \leq 0$ and $\mu_H(x;c) - \mu_H(y;c) \geq 0$ for every scale c. Conversely, if x S y does *not* hold, then there must be some k=1,...,n-1 for which $X_k - Y_k < 0$ and $k < m$,

or $X_k - Y_k > 0$ and $k \geq m$. If the first case applies, then there must exist scale $c'=[c'_1,...,c'_n]$ for which the gap $c'_{k+1} - c'_k$ is so large as to render the sum $2\sum_{j=1}^{m-1}\left(c'_{j+1} - c'_j\right)\left(Y_j - X_j\right)$ positive, in which case $s_L(x;c') < s_L(y;c')$ would be true. Similarly, if the second case applies, then there exists c' for which $c'_{k+1} - c'_k$ is so large that $2\sum_{j=m}^{n-1}\left(c'_{j+1} - c'_j\right)\left(Y_j - X_j\right)$ becomes negative, in which case $s_H(x;c') < s_H(y;c')$ is true. Either way, condition (iii) fails to hold, and (i) is equivalent to (iii).

(iii)↔(ii): Suppose that (iii) holds. Since $s(x;c) = s_L(x;c) + s_H(x;c)$ it follows immediately that (ii) holds as well. Conversely, suppose that (iii) does not hold, so that there is a c' such that $s_L(x;c') < s_L(y;c')$ or $s_H(x;c') < s_H(y;c')$ holds. If the first case applies, then it is clear that $s_L(x;c') - s_L(y;c') = s_L(x;c^\epsilon) - s_L(y;c^\epsilon)$, where c^ϵ is defined for $\epsilon > 0$ by $c^\epsilon_j = c'_j$ for $j < m$ and $c^\epsilon_j = c_{m-1} + \epsilon(j-m+1)$ for $j \geq m$. Moreover, for $\epsilon > 0$ small enough we obtain $s_L(x;c^\epsilon) - s_L(y;c^\epsilon) + s_H(x;c^\epsilon) - s_H(y;c^\epsilon) < 0$. Consequently, $s(x;c^\epsilon) < s(y;c^\epsilon)$, and so (ii) must not hold. A similar argument covers the second case where $s_H(x;c^\epsilon) < s_H(y;c^\epsilon)$, hence in either case (ii) will fail to hold.

Theorem III.5 Let x and y be any two distributions having the same median category. Then x FD y if and only if $s_L(x;c) \leq s_L(y;c)$ and $s_H(x;c) \geq s_H(y;c)$ for every scale c.

Proof

Let x and y have the same median category m. Suppose that x FD y, so that $Y_k - X_k \geq 0$ for all k. Then by (34) and (35) we have both $\mu_L(x;c) \geq \mu_L(y;c)$ and $\mu_H(x;c) \geq \mu_H(y;c)$ for every scale c. Since x and y have the same median category, it follows that $s_L(x;c) \leq s_L(y;c)$ and $s_H(x;c) \geq s_H(y;c)$ for every scale c. Conversely, if x FD y does not hold, then there must be some k=1,...,n-1 for which $Y_k - X_k < 0$. If $k < m$, then one can find a scale c′ for which $\mu_L(x;c') < \mu_L(y;c')$ and hence $s_L(x;c) > s_L(y;c)$. Alternatively, if $k \geq m$, then there is a scale c′ for which $\mu_H(x;c') < \mu_H(y;c')$ and hence $s_H(x;c') < s_H(y;c')$. Either way, the condition "$s_L(x;c) \leq s_L(y;c)$ and $s_H(x;c) \geq s_H(y;c)$ for every scale c" fails to hold, which completes the proof.

APPENDIX B

CHAPTER IV PROOFS

Theorems IV.1, IV.2, and IV.3 entail the application of second order dominance to demonstrate the robustness of mean rankings for various distributional ranges to concave transforms of the original scale, which are designed as follows

Definition There exists a function f defined over scale $c=[c_1,c_2,\ldots,c_n]$ such that $f(c_j)=c'_j$ for $j=1,2,\ldots,n$ and c' is a concave transform of c.

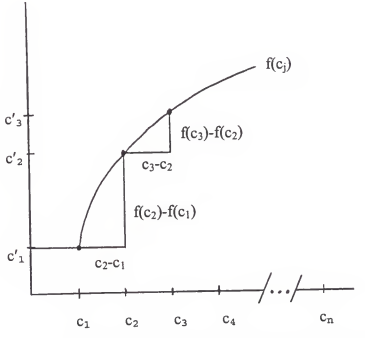


Figure 12. A concave transformation of reference scale c.

If c′ is a concave transform of c, then for all j=2,3,...,n-1

$\frac{f(c_{j+1})-f(c_j)}{c_{j+1}-c_j} \geq \frac{f(c_j)-f(c_{j-1})}{c_j-c_{j-1}}$, which implies that $\alpha_i \geq \alpha'_i$ for all

i=2,...,n-1, where $\alpha_i = \frac{c_{i+1}-c_i}{c_i-c_{i-1}}$ and $\alpha'_i = \frac{c'_{i+1}-c'_i}{c'_i-c'_{i-1}}$. Let L and H be

integers satisfying $1 \leq L < H \leq n$. Then c′ is a *concave transform* of c over the

categorical range i=L,L+1,...,H if $\alpha_i \geq \alpha'_i$ for all i=L+1,...,H-1.

Note that any c_i can be expressed in terms of c_L, c_{L+1} and $\alpha_{L+1},\alpha_{L+2},...,\alpha_i$ as follows

$$c_i = c_{i-1} + \frac{(c_i - c_{i-1})}{(c_{i-1} - c_{i-2})}(c_{i-1} - c_{i-2}) \tag{36}$$

$$c_i = c_{i-1} + \frac{(c_i - c_{i-1})}{(c_{i-1} - c_{i-2})}\frac{(c_{i-1} - c_{i-2})}{(c_{i-2} - c_{i-3})}\cdots\frac{(c_{L+2} - c_{L+1})}{(c_{L+1} - c_L)}(c_{L+1} - c_L) \tag{37}$$

$$c_i = c_{i-1} + (c_{L+1} - c_L)\prod_{j=L+1}^{i-1}\alpha_j \tag{38}$$

so that

$$c_i - c_{i-1} = (c_{L+1} - c_L)\prod_{j=L+1}^{i-1}\alpha_j . \tag{39}$$

Also note that if c′ is a concave transform of c over the categorical range i=L,L+1,...,H, then

$\alpha_i \geq \alpha'_i$ for all i=L,L+1,...,H, which implies that $\alpha_i\,\alpha_{i+1} \geq \alpha'_i\,\alpha'_{i+1}$ for all i=L,L+1,...,H, and

so on. Generally,

$$\prod_{j=p}^{q}\alpha_j \geq \prod_{j=p}^{q}\alpha'_j \text{ for integers p and q where } L+1 \leq p < q \leq H-1], \tag{40}$$

if c' is a concave transform of c over the categorical range i=L,L+1,...,H.

It is useful to first demonstrate the following basic result that may easily be applied to prove the three principle theorems

Lemma 1 Let x and y be any two distributions and let L and H be any two integers such that $1 \leq L < H \leq n$. One distribution, x, second order dominates another, y, for some scale c over the categorical range i=L,L+1,...,H, written

$$\sum_{j=L}^{k}(c_{j+1}-c_j)X_j \leq \sum_{j=L}^{k}(c_{j+1}-c_j)Y_j \text{ for all k=L,L+1,...,H} \tag{41}$$

only if x second order dominates y over the same categorical range for every scale c' that is a concave transform of c, written

$$\sum_{j=L}^{k}(c'_{j+1}-c'_j)X_j \leq \sum_{j=L}^{k}(c'_{j+1}-c'_j)Y_j \text{ for all k=L,L+1,...,H.} \tag{42}$$

Proof Let x and y be distributions and L and H integers such that x second order dominates y for scale c over the categorical range i=L,L+1,...,H

Show that (41) ⇒ (42): Using (39) to rewrite (41)

$$X_L(c_{L+1}-c_L)+\sum_{i=L+1}^{k}X_i(c_{L+1}-c_L)\prod_{j=L+1}^{i}\alpha_j \leq \tag{43}$$

$$Y_L(c_{L+1}-c_L)+\sum_{i=L+1}^{k}Y_i(c_{L+1}-c_L)\prod_{j=L+1}^{i}\alpha_j \quad \forall\, k \in [L,H].$$

Dividing (43) through by $(c_{L-1}-c_L)$ and re-arranging yields

$$Y_L - X_L \geq \sum_{i=L+1}^{k}\left(X_i - Y_i\right)\prod_{j=L+1}^{i}\alpha_j \quad \forall\, k \in [L,H]. \tag{44}$$

But by (40), $\prod_{j=L+1}^{i}\alpha_j \geq \prod_{j=L+1}^{i}\alpha'_j \;\forall$ i, since c′ is a concave transform of c, so

$$Y_L - X_L \geq \sum_{i=L+1}^{k}\left(X_i - Y_i\right)\prod_{j=L+1}^{i}\alpha'_j \quad \forall\, k \in [L,H]. \tag{45}$$

Multiplying through by $(c_{L+1}-c_L)$, applying (39) in reverse, and re-arranging yields

$$\sum_{j=L}^{k}(c'_{j+1}-c'_j)X_j \leq \sum_{j=L}^{k}(c'_{j+1}-c'_j)Y_j \;\forall\, k \in [L,H], \tag{46}$$

which is equivalent to (42). Q.E.D.

Theorem IV.1 One distribution, x, second order dominates another, y, for some scale c if and only if $\mu(x;c') \geq \mu(y;c')$ for every scale c′ that is a concave transform of c.

Proof Let x and y be distributions such that x SD(c) y. Applying Lemma 1 over the categorical range i=1,2,...,n

$$\sum_{i=1}^{n-1}(c'_{i+1}-c'_i)X_i \leq \sum_{i=1}^{n-1}(c'_{i+1}-c'_i)Y_i\,, \tag{47}$$

for any c′ that is a concave transform of c. Multiplying both sides of (IV.A9) by -1, adding $c_1 + \sum_{i=1}^{n-1}(c'_{i+1}-c'_i)X_i$ to both sides, and combining

terms yields

$$c'_1 + \sum_{i=1}^{n-1}(c'_{i+1} - c'_i)(1 - X_i) \geq c'_1 + \sum_{i=1}^{n-1}(c'_{i+1} - c'_i)(1 - Y_i) \tag{48}$$

for any c′ that is a concave transform of c. But by definition of the mean in (39), this implies that $\mu(x;c') \geq \mu(y;c')$ for every scale c′ that is a concave transform of c. This completes the proof in one direction.

Now assume that $\mu(x;c') < \mu(y;c')$ for some c′ that is a concave transform of c. Then, applying the definition of the mean in (39)

$$c'_1 + \sum_{i=1}^{n-1}(c'_{i+1} - c'_i)(1 - X_i) < c'_1 + \sum_{i=1}^{n-1}(c'_{i+1} - c'_i)(1 - Y_i) \tag{49}$$

for some c′. Subtracting $c_1 + \sum_{i=1}^{n-1}(c'_{i+1} - c'_i)X_i$ from both sides of (49) and multiplying by -1 yields

$$\sum_{i=1}^{n-1}(c'_{i+1} - c'_i)X_i > \sum_{i=1}^{n-1}(c'_{i+1} - c'_i)Y_i\,, \tag{50}$$

for some c′, which by Lemma 1 implies that

$$\sum_{i=1}^{s}(c_{i+1} - c_i)X_i > \sum_{i=1}^{s}(c_{i+1} - c_i)Y_i \tag{51}$$

for some category s=1,2,...,n-1, thus violating the definition of second order dominance. Hence, x ~SD(c) y. Q.E.D.

Theorem IV.2 Given distributions x and y with a common median, x second order dominates y up to the median for scale c if and only if $s_L(x;c') \leq s_L(y;c')$ for every scale c′ that is a concave transform of c.

Proof Let x and y be distributions with a common median category m such that x SD(c) y over the categorical range i=1,2,...,m. Then, applying Lemma 1

$$\sum_{j=1}^{k}(c'_{j+1}-c'_j)X_j \le \sum_{j=1}^{k}(c'_{j+1}-c'_j)Y_j \tag{52}$$

for all k=1,2,...,m-1 and for all concave transforms c′ of scale c. Hence

$$\sum_{j=1}^{m-1}(c'_{j+1}-c'_j)X_j \le \sum_{j=1}^{m-1}(c'_{j+1}-c'_j)Y_j \tag{53}$$

for all scales c′ that are concave transforms of c. Multiplying (53) through by 2 and adding $c'_1+\sum_{j=1}^{m-1}(c'_{j+1}-c'_j)$ to both sides yields

$$c'_1+\sum_{j=1}^{m-1}(c'_{j+1}-c'_j)(1-2Y_j) \le c'_1+\sum_{j=1}^{m-1}(c'_{j+1}-c'_j)(1-2X_j) \tag{54}$$

for all concave transforms c′ of scale c. Applying the definition of a lower mean in (32), this implies that

$$\mu_L(y;c') \le \mu_L(x;c') \tag{55}$$

for all concave transforms c′ of scale c. Multiplying both sides by -1 and adding c_m, the scale value for the median category, yields

$$c_m - \mu_L(x;c') \le c_m - \mu_L(y;c') \tag{56}$$

for all concave transforms c′ of scale c. Hence, by definition of s_L

$$s_L(x;c') \le s_L(y;c') \tag{57}$$

for every scale c′ that is a concave transform of c. This completes the proof in one direction.

Now, if $s_L(x;c') > s_L(y;c')$ for some c′, then

$$c_m - \mu_L(x;c') > c_m - \mu_L(y;c') \tag{58}$$

for some c′, so that

$$\mu_L(x;c') < \mu_L(y;c') \tag{59}$$

for some c′. Applying the definition of a lower mean from (32) yields

$$c'_1 + \sum_{j=1}^{m-1}(c'_{j+1} - c'_j)(1 - 2X_j) < c'_1 + \sum_{j=1}^{m-1}(c'_{j+1} - c'_j)(1 - 2Y_j) \tag{60}$$

for some c′. But then

$$\sum_{j=1}^{k}(c'_{j+1} - c'_j)Y_j < \sum_{j=1}^{k}(c'_{j+1} - c'_j)X_j \tag{61}$$

for some c′, which violates the definition of second order dominance below the median in (27), so that x cannot second order dominate y up to the median for scale c′. By Lemma 1 it must also be the case that x ~SD(c) y over the categorical range j=1,2,...,m, given that c′ is a concave transform of c. Q.E.D.

Theorem IV.3 Given two distributions x and y with a common median, y second order dominates x above the median for scale c if and only if $s_H(x;c') \leq s_H(y;c')$ y for every scale c′ that is a concave transform of c.

Proof Let x and y be distributions with a common median category m such that x SD(c) y above the median, then by Lemma 1

$$\sum_{j=m}^{k}(c'_{j+1} - c'_j)Y_j \leq \sum_{j=m}^{k}(c'_{j+1} - c'_j)X_j \tag{62}$$

for all k=m,m+1,...,n-1 and for all concave transforms c′ of scale c. Hence

$$\sum_{j=m}^{n-1}(c'_{j+1}-c'_j)Y_j \le \sum_{j=m}^{n-1}(c'_{j+1}-c'_j)X_j \tag{63}$$

for all concave transforms c′ of scale c. Multiplying through (63) by -2 and adding $c'_m + \sum_{j=m}^{n-1}(c'_{j+1}-c'_j)$ to both sides yields

$$c'_m + \sum_{j=m}^{n-1}(c'_{j+1}-c'_j)(1-2Y_j) \ge c'_m + \sum_{j=m}^{n-1}(c'_{j+1}-c'_j)(1-2X_j) \tag{64}$$

for all concave transforms c′ of scale c. Applying the definition of an upper mean in (33), this implies that

$$\mu_H(y;c') \ge \mu_H(x;c') \tag{65}$$

for all concave transforms c′ of scale c. Subtracting c'_m from both sides yields

$$\mu_H(y;c') - c'_m \ge \mu_H(x;c') - c'_m \tag{66}$$

for all concave transforms c′ of scale c. Hence, by definition of s_H

$$s_H(y;c') \ge s_H(x;c') \tag{67}$$

for every scale c′ that is a concave transform of c. This completes the proof in one direction.

Now, if $s_H(y;c') < s_H(x;c')$ for some c′, then

$$\mu_H(y;c') - c'_m < \mu_H(x;c') - c'_m \tag{68}$$

for some c′, so that

$$\mu_H(y;c') < \mu_H(x;c') \tag{69}$$

for some c′. Applying the definition of an upper mean from (33) yields

$$c'_m + \sum_{j=m}^{n-1}(c'_{j+1}-c'_j)(1-2Y_j) < c'_m + \sum_{j=m}^{n-1}(c'_{j+1}-c'_j)(1-2X_j) \tag{70}$$

for some c′. But then

$$\sum_{j=m}^{n-1}(c'_{j+1}-c'_j)Y_j > \sum_{j=m}^{n-1}(c'_{j+1}-c'_j)X_j \tag{71}$$

for some c′, which violates the definition of second order dominance above the median in (28), so that y cannot second order dominate x above the median for scale c′. By Lemma 1 it must also be the case that y ~SD(c) x over the categorical range j=m,m+1,...,n, given that c′ is a concave transform of c. Q.E.D.

BIBLIOGRAPHY

Anand, Sudhir and Vladimir Shkolnikov, eds. Measuring Health Globally. Oxford University Press, forthcoming.

Anderson, Kathryn H., and Richard V. Burkhauser. "The Retirement-health nexus: a new measure of an old puzzle," *Journal of Human Resources* 20 (1985), pp. 315-330.

Atkinson, Anthony B., and John Micklewright. The economic transformation in Eastern Europe and the distribution of income. Cambridge University Press, 1992.

Appels, A., H. Bosma, V. Grabauskas, A Gostautas, and F. Sturmans. "Self-reported health and mortality in a Lithuanian and a Dutch population," *Social Science and Medicine* 42:5 (1996), pp. 681-689.

Backlund, Eric, Paul D. Sorlie, and Norman J. Johnson. "The shape of the relationship between income and mortality in the United States: evidence from the National Longitudinal Mortality Study," *Annals of Epidemiology* 6:1(1996), pp. 12-20.

Black, Douglas, J. N. Morris, Cyril Smith, and Peter Townsend. Inequalities in Health: The Black Report. Penguin Books, New York, 1982.

Blumberg, Linda, John Holahan, and Marilyn Moon. "Options for reforming the Medicaid matching formula," The Urban Institute, Washington, D.C., February 18, 1993.

Butler, J.S., Richard V. Burkhauser, Jean M. Mitchell, and Theodore P. Pincus. "Measurement error in self-reported health variables," *The Review of Economics and Statistics* (1987), pp. 645-650.

Culyer, A.J., and Adam Wagstaff. "Equity and equality in health and health care," *Journal of Health Economics* 12 (1993), pp. 431-457.

Deaton, Angus, and Christina Paxson. "Health, income, and inequality over the life-cycle," Research Program in Development Studies, Princeton University. Draft as of July 26, 1997.

_____ "Aging and inequality in income and health," *American Economic Review* 88(2): May, 1998, pp. 248-253.

Dwyer, Debra Sabatini and Olivia S. Mitchell. "Health problems as determinants of retirement: are self-rated measures endogenous?" *Journal of Health Economics* 18 (1999), pp. 173-193.

Foster, James E. and Amartya Sen. "On Economic Inequality After a Quarter of a Century," in Sen (1997).

Foster, James E. and A. F. Shorrocks. "Poverty orderings," *Econometrica* 56 (1988).

Foster, James E. and Michael Wolfson. "Polarization and the decline of the middle class," mimeo, Vanderbilt University, 1993.

_____ "Measuring the middle," mimeo, Vanderbilt University, 1993.

General Accounting Office. "Medicaid: Alternatives for improving the distribution of funds to states," GAO/HRD-91-66FS, Washington, D.C., May 20, 1991

_____ "Medicaid: Alternatives for improving the distribution of funds to states," GAO/HRD-93-112FS, Washington, D.C., August 20, 1993.

Gerdtham, U. -G., M. Johannesson, L. Lundberg, and D. Isacson. "A note on validating Wagstaff and van Doorslaer's health measure in the analysis of inequalities in health," *Journal of Health Economics* 18(1999), pp. 117-124.

Grant, Mark D., Zdzisiaw H. Piotrowski and Rick Chappell. "Self-Reported Health and Survival in the Longitudinal Study of Aging, 1984-1986," *Journal of Clinical Epidemiology* 48:3 (1995), pp. 375-387.

Grossman, Michael J. "On the concept of health capital and the demand for health," *Journal of Political Economy* 80:2 (1972), pp. 223-255.

Hanoch, G. and H. Levy. "The efficiency analysis of choices involving risk," *Review of Economic Studies* 36 (1969), pp. 335-346.

Hadar, Joseph and William R. Russell. "Rules for ordering uncertain prospects," *American Economic Review* 59(1969), pp. 25-34.

House, James S., Ronald C. Kessler, and A. Regula Herzog. "Age, socioeconomic status, and health," *The Milbank Quarterly* 68:3 (1990), pp. 383-411.

Hurd, Michael D., and Kathleen McGarry. "Evaluation of the subjective probabilities of survival in the health and retirement study." *The Journal of Human Resources* XXX(1995S), pp. S268-S292.

Idler, Ellen, and Ronald Angel. "Self-rated health and mortality in the NHANES-I Epidemiologic follow-up study," *American Journal of Public Health* 80:4 (1990), 446-452.

Idler, Ellen, and Stanislav Kasl. "Health perceptions and survival: do global evaluations of health status really predict mortality?" *Journal of Gerontology: Social Sciences* 46:2(1991), pp. S55-65.

_____ "Self-ratings of health: do they also predict change in functional ability?" *Journal of Gerontology: Social Sciences* 50B:6 (1995), pp. S344-353.

Illsey, Raymond and Julian Le Grand. "The Measurement of Inequality in Health," in Health and Economics: Proceedings of Section F (Economics) of the British Association for the Advancement of Science, 1986. McMillan Press, 1987.

Jencks, Christopher and Barbara Boyle Torrey. "Beyond income and poverty: trends in social welfare among children and the elderly since 1960," in The Vulnerable, John L. Palmer, Timothy Smeeding, and Barbara Boyle Torrey, eds. The Urban Institute Press, Washington, D.C., 1988, pp. 229-273.

Kaplan, George A., Teresa Seeman, Richard D. Cohen, Lisa P. Knudsen, and Jack Guralnik. "Mortality among the elderly in the Alameda County Study: behavioral and demographic risk factors," *American Journal of Public Health* 77:3 (1987), pp. 307-312.

Kakwani, Nanak C. "Measurement of tax progressivity: an international comparison," *Economic Journal* 87(1977), pp. 71-80.

Kakwani, Nanak C., Adam Wagstaff and Eddy van Doorslaer. "Socioeconomic inequalities in health: measurement, computation, and statistical inference," *Journal of Econometrics* 77(1997), pp. 87-103.

Kawachi, Ichiro and Bruce P. Kennedy. "The relationship of income inequality to mortality: does the choice of indicator matter?" *Social Science and Medicine* 45:7 (1997), pp. 1121-1127.

Kawachi, Ichiro, Bruce P. Kennedy, Kimberly Lochner and Deborah Prothrow-Stith. "Social capital, income inequality, and mortality," *American Journal of Public Health* 87:9 (1997), pp. 1491-1498.

Kennedy, Bruce P., Ichiro Kawachi, Roberta Glass and Deborah Prothrow-Stith. "Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis," *British Medical Journal* 317 (1998), pp. 917-921.

Kennedy, Bruce P., Ichiro Kawachi, and Deborah Prothrow-Stith. "Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States," *British Medical Journal* 312 (1996), pp. 1004-1007.

Kunst, Anton. "Overviews of the state of the art with respect to measuring socio-economic indicators and their association to health indicators," in Measuring Health Inequality Globally, Sudhir Anand, Finn Diderichsen, and Vladimir Shkolnikov, eds. Oxford University Press (forthcoming in 1999).

Kunst, Anton, Jose Geurts, and Japp van den Berg. "International variation in socioeconomic inequalities in self-reported health," *Journal of Epidemiology and Community Health* 49(1995), pp. 117-123.

Le Grand, Julian, and Matthew Rabin. "Trends in British health inequality, 1931-1983," in Public and Private Health Services: Complimentarities and Conflicts, A.J. Culyer and Bengt Jeonsson, eds. Oxford Press, New York, 1986.

Livingston, Helen E, Legislative Reference Service, in *Social Security Amendments of 1955: Hearings before the Committee on Finance, United States Senate.* Library of Congress. 1956, p. 73.

Manning, Willard G., Jr., Joseph P. Newhouse and John E. Ware. "The status of health in demand estimation; or, beyond excellent, good, fair, and poor," in Economic Aspects of Health, Victor R. Fuchs, ed. The University of Chicago Press, Chicago (1982).

Marmot, Michael, Amanda Feeney, Martin Shipley, Fiona North, and S.L. Syme. "Sickness absence as a measure of health status and functioning: from the UK Whitehall II study," *Journal of Epidemiology and Community Health* 49 (1995), pp. 124-130.

McDonough, Peggy, Greg J. Duncan, David Williams, and James House (1997). "Income dynamics and adult mortality in the United States, 1972 through 1989," *American Journal of Public Health* 87:9 (1997), pp. 1476-1483.

Meredith, Lisa S., and Albert L. Siu. "Variation and quality of self-reported health data: Asians and Pacific Islanders compared with other ethnic groups," *Medical Care* 33:11 (1995), pp. 1120-1131.

Mossey, Jana, and Evelyn Shapiro. "Self-rated health: a predictor of mortality among the elderly," *American Journal of Public Health* 72:8 (1982), 800-808.

Muurinen, Jaana-Marja and Julian Le Grand. "The economic analysis of inequalities in health," *Social Science and Medicine* 20:10 (1985), pp. 1029-1035.

Mossey, Jana, and Evelyn Shapiro. "Self-rated health: a predictor of mortality among the elderly," *American Journal of Public Health* 72:8 (1982), 800-808.

Panis, Constantijn W.A., and Lee A. Lillard. "Health inputs and child mortality: Malaysia," *Journal of Health Economics* 13 (1994), pp. 455-489.

Power, Chris, Sharon Matthews, and Orly Manor. "Inequalities in self-rated health: explanations from different stages in life," *The Lancet* 351(April 4, 1998), 1009-1014.

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. Securing Access to Health Care, Volume 1: Report. Government Printing Office, Washington, D.C., 1983.

Ruhm, Christopher J. "Are recessions good for your health?" National Bureau of Economic Research working paper no. 5570. Cambridge, Massachusetts. May, 1996.

Sen, Amartya. "Poverty: an ordinal approach to measurement," *Econometrica* 44 (1976).

_____ Inequality Reexamined, Harvard University Press, Cambridge, 1992.

_____ On Economic Inequality, Enlarged Edition, Clarendon Press, Oxford, 1997.

Singer, Eleanor, Robin Garfinkel, Steven Cohen, and Leo Srole. "Mortality and mental health: evidence from the Midtown Manhattan restudy," *Social Science and Medicine* 10 (1976), 517-525.

Strauss, John, and Duncan Thomas. "Health, nutrition, and economic development," *Journal of Economic Literature* 36 (1998), pp. 766-817.

Tobin, James. "On limiting the domain of inequality." *The Journal of Law and Economics* 13(2): October, 1970, 263-277.

Van Doorslaer, Eddy, Adam Wagstaff, Han Bleichrodt, Samuel Calonge, Ulf-G. Gerdtham, Michael Gerfin, Jose Geurts, Lorna Gross, Unto Hakkinen, Robert E. Leu, Owen O'Donnell, Carol Propper, Frank Puffer, Marisol Rodriquez, Gun Sundberg, and Olaf Winkelhake. "Income-related inequalities in health: some international comparisons." *Journal of Health Economics* 16 (1997), pp. 93-112.

Van Vliet, Rene C.J.A., and Bernard M.S. Van Praag. "Health status estimation on the basis of MIMIC-health care models," *Journal of Health Economics* 6 (1987), pp. 27-42.

Waidmann, Timothy, John Bound, and Michael Schoenbaum. "The illusion of failure: trends in self-reported health of the U.S. elderly," *The Milbank Quarterly* 73:2 (1995), pp. 253-285.

Wagstaff, Adam."The demand for health: some new empirical evidence," *Journal of Health Economics* 5 (1986), pp. 195-233.

Wagstaff, Adam. "The demand for health: theory and applications," *Journal of Epidemiology and Community Health* 40 (1986), pp. 1-11.

Wagstaff, Adam, Pierella Paci and Eddy van Doorslaer. "On the measurement of inequalities in health," *Social Science and Medicine* 33:5 (1991), pp. 545-557.

Wagstaff, Adam, Eddy Van Doorslaer, and Pierella Paci. "Equity in the finance and delivery of health care: some tentative cross-country comparisons," *Oxford Review of Economic Policy* 5:1 (1989), pp. 89-112.

Wagstaff, Adam, and Eddy Van Doorslaer. "Measuring inequalities in health in the presence of multiple-category morbidity indicators," *Health Economics* 3 (1994), pp. 281-291.

_____ "Equity in Health Care Finance and Delivery." prepared for chapter 40, North Holland Handbook of health Economics (forthcoming), A.J. Culyer and J.P. Newhouse, eds. October, 1998.

Wagstaff, Adam, and Eddy Van Doorslaer. "Inequalities in health: methods and results for Jamaica," prepared for the Human Development Department of the World Bank, first version: February 1998.

Woolson, Robert F. Statistical Methods for The Analysis of Biomedical Data, John Wiley & Sons, New York, 1987.

CMS LIBRARY
3 8095 00005989 5
c.1